VISIT…

LANZAROTE Caliente.com

序　言

经过几年的艰辛努力，我们终于能够把这本关于现代西方语言学研究方法的书呈献给我国有志于从事语言学与应用语言学研究的人们，不无欣慰。

"工欲善其事，必先利其器。"这是我国的古训。庄子在《庖丁解牛》里讲顺应自然规律的重要性，但是解牛离不开刀，"彼节者有间，而刀刃者无厚，以无厚入有间，恢恢乎其于游刃必有余地矣。"可见掌握自然规律离不开工具。古希腊亚里士多德作《工具论》，阐述他所使用的研究方法（三段论演绎法），培根在1620著《新工具论》，对研究方法作了重要的补充（归纳法）。培根是近代实验方法的倡导者，但忽视了数学方法；笛卡儿推崇数学方法，他自述说，"从青年时代以来，就发现了某些途径，引导我作了一些思考，获得一些公理，我从这些思考和公理形成一种方法，凭借这种方法，我觉得自己有了依靠，可以逐步增进我的知识，并且一点一点把它提高到我的平庸才智和短促生命所能达到的最高点。"① 天文学家伽利略被认为是把实验方法和数学方法巧妙地结合起来的一位代表，而伽利略的成就恰恰是得益于望远镜的发明。库恩（Kuhn）提出科学知识的发展遵循"前科学—常规科学—危机—科学革命—新的常规科学"的范式，体现了生产力发展的新方法（包括新工具）应该是科学革命的重要催生剂。我们很难想像没有现代的微观技术而能够进行基因的研究，没有星际飞行器而能够观察宇宙空间。

语言学研究也要讲究方法，早在1937年胡朴安在其《中国训诂学史》里就指出，像《尔雅》那样的典籍，只能说是训诂书，而不能称之为训诂学，"凡称为学，必有学术上的方法。训诂之方法，至清朝汉学家，始能有条理有系统之发现。"胡当时就指出，"训诂学方法之新趋势，惟有甲骨今文之考证，与统计学的推测，二法而已。"他自己就统计了《论语》中"仁"字出现的概率及其意义，最后"惟余确信训诂在统计上极有价值之方法，惟用之者须有旧式训诂学根基。"②现代西方语言学的研究方法更为广博，因为它和很多语言学交叉学科的发展分不开：理论语言学使用了数理逻辑的方法、应用语言学使用了教育测量和统计的方法、心理语言学使用了心理测量的方法、社会语言学（包括文化语言学）使用了社会学的调查方法、计算语言学使用了计算机的方法、神经语言学使用了神经生理和解剖学的方法等等，不一而足。所以语言学工作者都面临着一个重新学习的挑战。

我们曾经对最近三年的四本外语研究期刊中的755篇文章作过一次小统计，这四本期刊是：

- 《外语教学与研究》
- 《外国语》
- 《外语界》
- 《现代外语》

调查分三个方面，一个是研究领域，一个是研究方法类型，一个是数据类型。

---

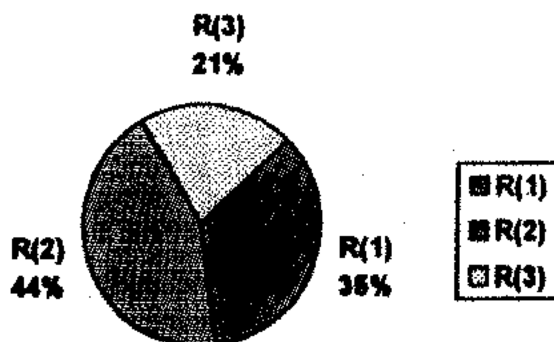
① 《16—18世纪西欧各国哲学》，三联书店1958年版，第104页。

② 胡朴安《中国训诂学史》，中国书店1983年根据商务1937年版影印，第3页，第357—359页。

研究领域分为三类：应用语言学（R1）；与语言学相关学科（R2），如社会语言学、心理语言学、语用学等；理论语言学（R3），包括句法学、语音学、语义学等等。结果是：

表一　　研究领域

	R（1）	R（2）	R（3）	总计
93	89	99	63	251
94	99	109	50	258
95	79	123	44	246
总计	267	331	157	755

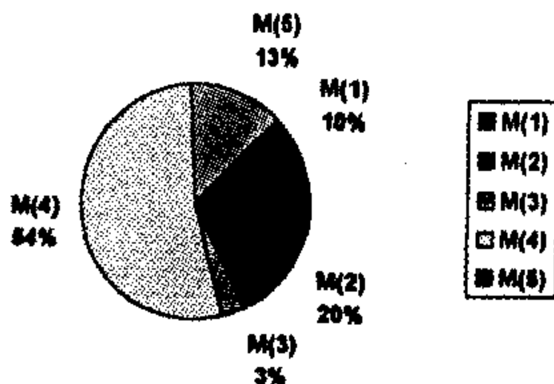


图一　研究领域

研究方法类型分为5种：理论性（M1），描述性（M2），实验性（M3），思辨性（M4）（指对一些问题进行讨论，随意举些例证），介绍性（M5）。

表二　　研究方法类型

	M(1)	M(2)	M(3)	M(4)	M(5)	总计
93	26	56	7	133	29	251
94	25	48	9	137	39	258
95	24	48	8	133	33	246
总计	75	152	24	403	101	755

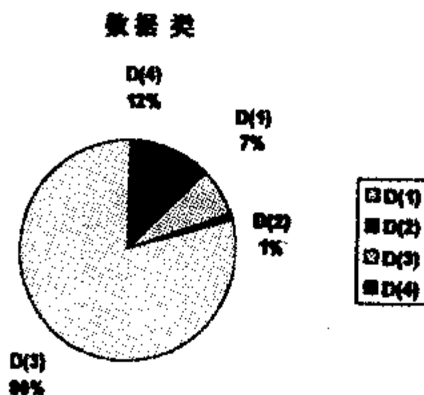


图二　研究方法类型

数据类型分为4种：统计数据（D1），指用统计方法处理过的数据；计算机数据（D2）；不依赖数据（D3），只随意举几个例证；数据罗列（D4），无统计数字。

表三　　数据类型

	D(1)	D(2)	D(3)	D(4)	总计
93	17	3	197	34	251
94	26	0	203	29	258
95	11	3	203	29	246
总计	54	6	603	92	755



图三　数据类

这几个统计数字说明，我国的外语工作者对语言学的各个领域都有兴趣，但是却没有足

够的方法支持他们的研究，54%的人使用的是简单的思辩性的方法，随机性很大。这些研究方法的成果难登大雅之堂。这倒不是说思辩性研究不可取，对母语的研究，也可依赖我们对母语的直觉观察来进行思考，提出理论模型。但是对非母语的描述性研究和实验性研究却必须以数据为依归，而我们有80%的研究都是“不依赖数据”的。

我们的调查只限于几种主要的外语刊物，相信再扩大调查范围，其结果也会是相同的。

究其原因，最根本的一条是我国外语工作者和研究者缺乏语言学研究方法的必要的训练，或是原来所掌握的方法已经老化，未能更新。随着语言学研究范围的伸延，这个问题变得更为严重。在国外，有识之士已经在大学开设研究方法的课程，甚至还建立了研究方法系。而我国很多大学的研究生课程还没有开设研究方法及其有关的（如统计学）课程。我们这本书正是针对这种需要编写的，希望能解燃眉之急。

美国语言学家Nida在我国讲学期间曾经指出，语言学研究在中国是一块未开垦的处女地。他对我国的语言学研究的传统当然并不熟悉，指的是用现代语言学研究方法来研究语言。无论从理论语言学，还是从应用语言学、社会语言学、心理语言学、语用学、计算语言学的角度看，只要我们肯挥锄破土，都会有所收获。我国广大的外语工作者都有此愿望，可惜的是他们也缺乏必要的工具，所以这本书也是为他们而编写的。

我们在编写本书时首先要考虑的问题的是语言学边界不断扩充，语言学研究方法再不能固守它原来的疆土。但是它的范围实在太大，而且我们的学识浅陋，必须有所取舍。所以我们必须定出书的范围：一个是“当代的”，像语文学、历史语言学就只能割爱。二是“西方的”，像谈论中国的考据学、方言学的方法论的书已有不少，无须我们去班门弄斧；谈“西方的”绝非崇洋媚外，而是希望引进一些新的方法和手段，丰富我们的语言学研究，使它能够和国际接轨。三是只介绍我们所熟悉的，像神经语言学的研究方法非常专门，非我们所长，只能有待来者。四是受篇幅所限，很多研究方法只能浮光掠影，还不能深入讨论。

本书分为三篇：理论方法篇、描写方法篇、实验方法篇。这样区分是为了讨论问题的方便：第一篇是针对理论语言学的方法论的；第二篇是针对从人类学语言学传统开始，一直发展为描写语言学、社会语言学、文化语言学、篇章语言学的方法论的；第三篇是针对应用语言学（教育语言学）、心理语言学、认知语言学的方法论的。这是以方法论为纲来展开讨论，实际上很多语言学研究往往会交叉使用不同的方法。

我们还想就理论和方法的关系说几句话：

一、理论和方法的一致性。理论和方法首先是不可分的：以Chomsky为代表的理论语言学潮流的目标是认识人类语言的普遍现象（共项），建立语言的形式化的表述系统，一个包括初始元的公理系统，就不得不使用现有公理系统（如数学、几何、形式逻辑等）中的初始元，因此需要用到数学方法和逻辑方法。人类语言学和社会语言学以研究语言的差异为目标，就不得不使用许多语言调查和语言描写的方法。心理语言学以了解人类习得和使用语言的心理过程为目标，而心理过程是看不见、摸不着的，就不得不使用实验方法和统计方法。计算语言学在进行自然语言处理时，有不同的理论和相应的方法：一种观点是让计算机具有和人相同的广泛的知识和逻辑推理的能力，这需要用演绎逻辑的手段把规则形式化，然后处理自然语言，这是人工智能的方法。另一种观点是把庞大的语料库放到计算机里，使用定量分析方

法，对自然语言现象进行概率性的预测。这是语料库方法或概率的方法。

正因为理论和方法是不可分的，所以我们在介绍方法时也要接触理论，甚至理论的发展，如讨论语言描写的方法时，不得不概括从人类语言学到社会语言学和文化语言学的过程。但是限于篇幅，我们不可能对语言理论和语言史谈得很详尽。

二、上面谈到，现代语言学的一个鲜明的特征是学科的交叉性，我们要在方法论上打好基础，还不得不涉猎一些相关学科的理论和方法。在理论语言学方面，必须学一点数学和逻辑学；在社会语言学方面，必须学一点社会学；在心理语言学方面，必须学一点心理学；在计算语言学方面，必须学一点计算机科学。而在方法论本身，还必须懂得定性方法和定量方法。对工具的掌握和使用应该持动态的观点，这是因为工具的发展和科学与生产力的发展密切相关，是最活跃的因素。和任何科学的探究一样，工具的改善和发展是一个不会终结的过程；而怎样巧用工具还要我们自己不断揣摩。随着各个语言学科的深入发展，我们的研究方法上必须不断更新。在本书里，我们也尽量反映一些新观点、新方法、新资料，有些材料还是直接从国际互联网下载的。但是我们所谈的只是一些基本的原则和方法，要升堂入室，还要靠读者自己努力。

三、虽然理论与方法是一致的，但方法之间却是互补的。对某些语言现象的理论上的思考可以辅以语料的调查，甚至语言使用的心理实验研究。对语言现象的描写又可以上升为形式化的规则。定量的分析依靠的是数据，而定性分析所依赖的是词语，两者相辅相成，可以使研究更为有血有肉。使用不同的方法来探究同一语言现象可以互相引证，提高观察的信度和效度。有些研究的目标虽然是一致的，但使用不同的方法可以有不同的角度，使观察更为全面，例如话语分析可以从定性方法的角度去研究，那就需要建立理论框架和描写框架，目的是分析词语本身。我们也可以从定量方法的角度去研究，那就要用到语料库方法和统计方法，目的是了解文体的统计学特征。我们还可以从人工智能的角度去研究，那就要用到实验方法和计算机方法，目的是进行自然语言处理。这些方法都很不一样，所以我们在本书里，只好在三个不同的地方来讨论这个问题。工具本身随着科学和生产水平的提高也会越来越精细，就好像显微镜一样，可以不断增加倍数，所以我们需要从不同的角度去观察现象，甚至观察所使用的工具本身。最近兴起的多倾向多方法的设计（Multitrait-Multimethod Design，MT-MM）以及与之配合的结构方程模型[①]虽然和实验方法和统计方法有关，但其含义却是适用于整个语言学研究。

本书第一篇由宁春岩负责，第二篇和第三篇由桂诗春负责。正如我们在上面所谈到的，不同的语言学科的研究对象和目标不一样，所采取的观点和方法也不一样，因此几个部分之间的看法和说法可能有抵牾之处，我们也不求其统一，希望读者留意。全书的涉及范围很广，错误疏漏之处在所难免，敬请海内外名家和广大读者批评、指正。我们感到遗憾的是我国的语言学研究有悠久的传统，近年来国内外的学者在汉语研究上又取得丰硕成果，这都未能在书中反映，一是我们没有这个能力，二是没有这么多的篇幅。如果有哪一位学者写一本关于汉语的语言学方法论，我们相信那一定是更为精彩的鸿篇巨著。

这本书是国家教委人文、社会科学研究“八五”规划的重点项目，得到了国家教委的资

---

① 请参看本书10.7.2.4.和11.7.7.2.

助；北京外国语大学的外语教学与研究出版社全力支持，使它能够短期内面世。我们在此一并表示衷心的感谢！

编著者

1996年9月

于广州白云山下

目　录

序言 …………………………………………………………………………………… (Ⅰ)

上篇　理论方法篇

1．西方现代理论语言学的一般理论特征 ……………………………………… (3)
1.1. 西方现代理论语言学的认识论基础 ……………………………………… (3)
1.2. 西方现代理论语言学的理论对象 ………………………………………… (5)
1.3. 西方现代理论语言学的理论目标 ………………………………………… (6)
1.4. 西方现代理论语言学的理论表述特征 …………………………………… (7)

2．西方现代理论语言学理论方法的总体特征 ………………………………… (8)
2.1. 抽象理论模型法 …………………………………………………………… (8)
2.2. 模块式的理论模型 ………………………………………………………… (10)
2.3. 语言理论模型的外模块系统 ……………………………………………… (11)
2.4. 几个需要说明的问题 ……………………………………………………… (16)

3．语法系统的内模块结构 ……………………………………………………… (18)
3.1. 语法系统的总体任务 ……………………………………………………… (18)
3.2. 语法系统的内模块系统形式 ……………………………………………… (22)
3.3. 语法系统内模块的建立 …………………………………………………… (34)
3.4. 语法系统内模块建立的普遍语法原则 …………………………………… (37)

4．西方现代理论语言学方法论的具体特征 …………………………………… (42)
4.1. 对比研究法 ………………………………………………………………… (42)
4.1.1. 在可能与不可能之间寻找限制 ………………………………………… (42)
4.1.2. 对比研究法 ……………………………………………………………… (43)
4.2. 事实关联研究 ……………………………………………………………… (50)
4.3. 理论逼近法 ………………………………………………………………… (54)
4.4. 在语言直觉知识面前 ……………………………………………………… (60)
4.5. 比较研究法 ………………………………………………………………… (64)
4.6. 二元属性描写法 …………………………………………………………… (72)
4.7. 空范畴研究 ………………………………………………………………… (76)
4.8. 西方现代语言学与其他语言理论的关系 ………………………………… (79)

中篇　描写方法篇

5．语言学中的人类学传统 ………… (85)
5.1. 语言学在人类学中的地位 ………… (85)
5.2. 人类语言学的发展 ………… (85)
5.3. Malinowski 的贡献 ………… (86)
5.3.1. 语言环境 ………… (87)
5.3.2. 语言功能 ………… (87)
5.4. 法国的传统 ………… (88)
5.5. 美洲的传统 ………… (88)
5.5.1. Boas 的贡献 ………… (89)
5.5.2. 描写语言学和结构语言学的诞生 ………… (89)
5.6. 人类语言学的特征 ………… (90)
5.7. 人类语言学和定性研究方法 ………… (91)
5.7.1. 生态心理学 ………… (91)
5.7.2. 整体文化人类学 ………… (91)
5.7.3. 文化语言交际学 ………… (91)
5.7.4. 认知人类学 ………… (92)
5.7.5. 符号连接主义 ………… (92)

6．定性研究方法 ………… **(93)**
6.1. 定性研究方法的原则 ………… (93)
6.1.1. 自然观察 ………… (93)
6.1.1.1. 直接观察 ………… (93)
6.1.1.2. 参与性观察 ………… (93)
6.1.1.3. 个案研究 ………… (94)
6.1.2. 归纳分析 ………… (94)
6.1.3. 全面观 ………… (96)
6.1.4. 综合法 ………… (97)
6.1.5. 分类系统 ………… (98)
6.1.6. 动态的发展 ………… (100)
6.1.7. 抽样 ………… (101)
6.2. 进行定性研究的程序 ………… (102)
6.2.1. 定义所要描写的现象 ………… (102)
6.2.2. 使用定性方法收集数据 ………… (102)
6.2.3. 在数据中寻找型式 ………… (104)
6.2.4. 回到数据中去收集更多的数据以支持初步的结论 ………… (105)
6.2.5. 研究过程的循环往复 ………… (105)

6.3. 定性研究、描述性研究和实验性研究 …… (105)

7. 语言描写方法 …… (107)
7.1. 现代语言学的诞生 …… (107)
7.2. 语言描写的基本概念和模式 …… (108)
7.3. 语言描写的目标——抽象化 …… (110)
7.4. 语言要素的分布 …… (111)
7.4.1. 组合关系和聚合关系 …… (112)
7.4.2. 形式化 …… (113)
7.4.3. 一个实例 …… (114)
7.5. 语言形式的描写 …… (116)
7.5.1. 语音描写 …… (117)
7.5.2. 形态 …… (119)
7.5.3. 直接成分分析法 …… (120)
7.6. 型式分析 …… (122)
7.7. 话语分析 …… (123)
7.7.1. 话语分析的定性研究 …… (123)
7.7.1.1. 话语分析的理论模型 …… (123)
7.7.1.2. 话语分析的描写框架 …… (129)
7.7.2. 话语分析的定量研究 …… (133)
7.7.2.1. 统计学方法 …… (133)
7.7.2.2. 语料库方法 …… (138)

8. 社会语言学研究方法 …… (150)
8.1. 结构语言学和社会语言学 …… (150)
8.2. 社会语言学的抽样方法 …… (153)
8.2.1. 对样本的总体作出定义 …… (154)
8.2.2. 分层抽样 …… (154)
8.2.3. 样本数目 …… (156)
8.2.4. 判断抽样 …… (156)
8.2.5. 语言的抽样 …… (157)
8.3. 数据的收集 …… (157)
8.3.1. 数据的类型 …… (157)
8.3.1.1. 背景信息 …… (157)
8.3.1.2. 人工制品 …… (158)
8.3.1.3. 社会组织 …… (158)
8.3.1.4. 法律信息 …… (158)
8.3.1.5. 艺术数据 …… (158)
8.3.1.6. 普通常识 …… (158)

8.3.1.7. 语言使用的信念 …… (159)
8.3.1.8. 关于语码的数据 …… (159)
8.3.2. 数据收集程序 …… (159)
8.3.2.1. 内省法 …… (160)
8.3.2.2. 观察法 …… (160)
8.3.2.3. 访问 …… (161)
8.3.2.4. 文化语义学 …… (163)
8.3.2.5. 文化方法论与交互作用分析 …… (164)
8.4. 数据的描写和分析 …… (165)
8.4.1. 定性分析 …… (165)
8.4.2. 相关性研究 …… (169)
8.4.3. 交互作用分析 …… (173)
8.4.3.1. 交际的组成部分 …… (173)
8.4.3.2. 交际事件的分析 …… (179)
8.4.4. 数据的形式化描写 …… (185)
8.4.4.1. 变量规则 …… (185)
8.4.4.2. 语音 …… (186)
8.4.4.3. 句法 …… (187)
8.4.4.4. 语篇 …… (189)
8.4.5. 语言态度调查 …… (190)
8.4.5.1. 什么是语言态度调查? …… (190)
8.4.5.2. 语言态度调查方法 …… (190)
8.4.5.3. 语言态度调查的用途。 …… (193)
8.4.6. 语言用法调查 …… (194)
8.4.6.1. Fries 的《美国英语语法》…… (196)
8.4.6.2. Mittins 的英语用法调查 …… (197)
8.4.6.3. Quirk 等人的现代英语用法调查 …… (197)
8.4.7. 定量分析 …… (202)
8.4.7.1. 人口普查和调查 …… (202)
8.4.7.2. Greenberg-Lieberson 公式 …… (204)
8.4.7.3. Lieberson 的延伸 …… (206)
8.4.7.4. 定量方法在广义社会语言学中的应用 …… (207)

下篇　实验方法篇

9. 语言学研究的实验方法 …… (211)
9.1. 定量方法 …… (211)
9.2. 定性方法和定量方法 …… (212)
9.2.1. 操纵和控制 …… (213)

9.2.2. 逻辑实证主义 …… (213)
9.2.3. 演绎 …… (214)
9.2.4. 分析 …… (214)
9.2.5. 推断性 …… (215)
9.3. 实验方法的意义和应用 …… (216)
9.3.1. 实验方法的重要性 …… (216)
9.3.2. 实验方法在语言学的应用 …… (218)
9.3.2.1. 实验方法在应用语言学中的应用 …… (218)
9.3.2.2. 实验方法在心理语言学中的应用 …… (221)
9.4. 认知科学与实验方法 …… (223)
9.4.1. 认知科学的两条基本原则 …… (224)
9.4.1.1. 可计算性的原则 …… (224)
9.4.1.2. 信息处理的原则 …… (225)
9.4.2. 一般的实验方法 …… (229)
9.4.2.1. 表征类型分析法 …… (229)
9.4.2.2. 扣除法 …… (229)
9.4.2.3. 递加因素法 …… (230)
9.4.2.4. 双任务法 …… (231)
9.4.2.5. 信号监察法 …… (231)
9.4.2.6. 计算机模拟法 …… (232)
9.4.3. 实验方法类型 …… (233)
9.4.3.1. 潜伏性数据 …… (233)
9.4.3.2. 眼睛固视 …… (234)
9.4.3.3. 口头报告 …… (235)
9.4.3.4. 双耳实验 …… (237)
9.4.3.5. 辨认与回述 …… (238)
9.4.3.6. 判断 …… (239)
9.4.3.7. 转移 …… (239)
9.4.3.8. 概念学习实验 …… (240)
9.4.3.9. 在线测量 …… (240)
9.4.4. 话语分析的实验方法 …… (242)
9.4.4.1. 早期的研究 …… (242)
9.4.4.2. 语篇结构 …… (243)
9.4.4.3. 语篇结构的语言指示器 …… (244)
9.4.4.4. 短语层面现象 …… (245)
9.4.4.5. 计划的识别 …… (247)

10. 实验设计 …… (249)
10.1. 选择课题 …… (249)

10.1.1. 一般性的课题 …………………………………… (249)
10.1.2. 课题焦点 …………………………………… (251)
10.1.2.1. 可行性问题 …………………………………… (251)
10.1.2.2. 综合性还是分析性研究? …………………………………… (252)
10.1.2.3. 缩小课题范围 …………………………………… (252)
10.1.3. 决定目标 …………………………………… (254)
10.1.4. 形成研究计划或假设 …………………………………… (254)
10.2. 提出假设 …………………………………… (254)
10.2.1. 什么是假设? …………………………………… (254)
10.2.2. 观察和假设的关系 …………………………………… (255)
10.2.3. 假设是怎样来的? …………………………………… (255)
10.2.4. 建立备择假设 …………………………………… (256)
10.2.5. 在概念化基础上的假设 …………………………………… (256)
10.2.6. 检验假设 …………………………………… (257)
10.3. 文献评论 …………………………………… (258)
10.3.1. 评论的目的 …………………………………… (258)
10.3.1.1. 找出重要的变量 …………………………………… (258)
10.3.1.2. 明确研究方向 …………………………………… (258)
10.3.1.3. 综观全局 …………………………………… (258)
10.3.1.4. 决定意义和关系 …………………………………… (258)
10.3.2. 文献资料来源 …………………………………… (260)
10.3.2.1. 图书杂志 …………………………………… (260)
10.3.2.2. 电子资源 …………………………………… (260)
10.3.3. 文献资料检索 …………………………………… (263)
10.3.3.1. 选择兴趣领域和主题词 …………………………………… (263)
10.3.3.2. 检索有关题目和摘要 …………………………………… (264)
10.4. 决定变量 …………………………………… (264)
10.4.1. 自变量 …………………………………… (265)
10.4.2. 依变量 …………………………………… (265)
10.4.3. 自变量和依变量的关系 …………………………………… (265)
10.4.4. 调节变量 …………………………………… (266)
10.4.5. 控制变量 …………………………………… (267)
10.4.6. 介入变量 …………………………………… (267)
10.4.7. 变量的组合 …………………………………… (267)
10.4.8. 变量的操作定义 …………………………………… (269)
10.5. 操纵和控制变量 …………………………………… (270)
10.5.1. 影响内部效度的因素 …………………………………… (271)
10.5.2. 影响外部效度的因素 …………………………………… (272)
10.5.3. 选样的控制 …………………………………… (273)

10.5.4. 历史的控制 …… (274)
10.5.5. 选样和历史的综合控制 …… (276)
10.5.6. 工具的控制 …… (277)
10.6. 决定实验设计方案 …… (277)
10.6.1. 前实验设计 …… (277)
10.6.1.1. 一次性个案研究 …… (277)
10.6.1.2. 一组实验前后测试设计 …… (278)
10.6.1.3. 原组比较 …… (278)
10.6.2. 真正的实验设计 …… (278)
10.6.2.1. 只有实验后测试的控制组设计 …… (278)
10.6.2.2. 实验前后测试的控制组设计 …… (279)
10.6.3. 因子设计 …… (279)
10.6.4. 准实验设计 …… (281)
10.6.4.1. 时间次序设计 …… (281)
10.6.4.2. 等值时间样本设计 …… (281)
10.6.4.3. 非等值控制组设计 …… (282)
10.6.4.4. 分离样本的实验前后测试设计 …… (283)
10.6.5. 事后的实验设计 …… (283)
10.6.5.1. 相关研究 …… (283)
10.6.5.2. 标准组设计 …… (284)
10.7. 观察与测量程序 …… (284)
10.7.1. 测量的信度 …… (284)
10.7.1.1. 再测法 …… (285)
10.7.1.2. 平行试题法 …… (286)
10.7.1.3. 对半信度估算法 …… (287)
10.7.1.4. Kuder-Richardson 信度系数 …… (288)
10.7.1.5. α 系数 …… (288)
10.7.1.6. 阅卷员信度 …… (288)
10.7.1.7. 标准误 …… (289)
10.7.1.8. 影响信度的因素 …… (289)
10.7.1.9. 经典测试理论估算信度的问题 …… (290)
10.7.2. 测量的效度 …… (290)
10.7.2.1. 内容效度 …… (291)
10.7.2.2. 预测效度 …… (292)
10.7.2.3. 共时效度 …… (292)
10.7.2.4. 构想效度 …… (293)
10.7.3. 数据量表 …… (295)
10.7.3.1. 量表类型 …… (295)
10.7.3.2. 量表的建立 …… (297)

10.7.3.3. 语义微分分析 …… (299)
10.7.4. 记录观察结果的手段 …… (301)
10.7.4.1. 评分量表 …… (301)
10.7.4.2. 编码系统 …… (302)
10.7.5. 问卷与访问设计 …… (302)
10.7.5.1. 应该怎样提问? …… (303)
10.7.5.2. 应该怎样回答问题? …… (304)
10.7.5.3. 建立问卷或访问提纲 …… (307)

11. 统计方法 …… **(311)**
11.1. 描写统计方法 …… (312)
11.1.1. 数据的归纳 …… (312)
11.1.1.1. 列表 …… (312)
11.1.1.2. 频数表 …… (314)
11.1.2. 集中量 …… (315)
11.1.2.1. 平均数 …… (316)
11.1.2.2. 中位数 …… (316)
11.1.2.3. 众数 …… (316)
11.1.2.4. 几个集中量的比较 …… (316)
11.1.2.5. 修剪平均数 …… (317)
11.1.3. 离散量 …… (319)
11.1.3.1. 全距 …… (319)
11.1.3.2. 方差与标准差 …… (320)
11.1.3.3. 偏态值与峰值 …… (320)
11.1.4. 描写统计方法的计算机运算 …… (321)
11.2. 概率分布 …… (322)
11.2.1. 离散性概率分布 …… (323)
11.2.1.1. 随机变量和离散性概率分布 …… (323)
11.2.1.2. 二项分布 …… (324)
11.2.2. 连续性概率分布 …… (326)
11.2.2.1. 连续性随机变量 …… (326)
11.2.2.2. 正态分布 …… (326)
11.2.2.3. 标准分 …… (328)
11.2.3. 概率分布的计算机运算 …… (330)
11.3. 推断统计方法 …… (331)
11.3.1. 参数估计 …… (331)
11.3.1.1. 总体参数的点估计 …… (331)
11.3.1.2. 总体参数的区间估计 …… (331)
11.3.1.3. 比例的估计 …… (332)

11.4. 假设检验 …… (333)
11.4.1. 使用置信区间来检验假设 …… (333)
11.4.2. Z 检验 …… (333)
11.4.3. t 检验 …… (335)
11.4.3.1. 配对 t 检验 …… (337)
11.4.3.2. 独立样本的 t 检验 …… (338)
11.4.3.3. t 检验的计算机运算 …… (339)
11.4.4. χ^2 检验 …… (340)
11.4.4.1. 单向表的 χ^2 检验 …… (341)
11.4.4.2. 双向表的 χ^2 检验 …… (343)
11.4.4.3. χ^2 检验要注意的问题 …… (344)
11.4.4.4. χ^2 检验的计算机运算 …… (345)
11.5. 相关系数与线性回归 …… (346)
11.5.1. 协方差 …… (346)
11.5.2. 积差相关系数 …… (347)
11.5.3. 等级相关 …… (348)
11.5.4. 线性回归 …… (348)
11.5.5. 相关矩阵和部分相关 …… (351)
11.5.6. 相关系数的计算机运算 …… (351)
11.6. 方差分析 …… (354)
11.6.1. 单向方差分析 …… (354)
11.6.2. 重复测量设计 …… (356)
11.6.3. 双向方差分析 …… (357)
11.6.4. 固定效果和随机效果模型 …… (359)
11.6.5. 方差分析的计算机运算 …… (359)
11.7. 多元分析方法 …… (361)
11.7.1. 多元回归分析 …… (362)
11.7.2. 多元方差分析 …… (367)
11.7.3. 主要成分分析和因子分析 …… (369)
11.7.3.1. 相关系数的几何法 …… (369)
11.7.3.2. 形心法 …… (370)
11.7.3.3. 主要成分分析 …… (371)
11.7.3.4. 因子分析 …… (374)
11.7.4. 聚类分析 …… (376)
11.7.4.1. 距离矩阵 …… (376)
11.7.4.2. 分层聚类分析 …… (377)
11.7.5. 多维度量表 …… (380)
11.7.6. 判别分析 …… (381)
11.7.7. Lisrel 模型 …… (385)

11.7.7.1. 验证性因子分析 …………………………………………………… (386)
11.7.7.2. 结构方程模型 …………………………………………………… (389)
11.8. 抽样数目的估算 …………………………………………………… (392)
11.8.1. 决定样本数的几个因素 …………………………………………… (392)
11.8.2. 简单的随机抽样 …………………………………………………… (393)
11.8.3. 分层抽样 …………………………………………………………… (393)
11.8.4. 整群抽样和系统抽样 ……………………………………………… (394)
11.9. 小结………………………………………………………………… (395)
参考书目……………………………………………………………………… (398)
附录…………………………………………………………………………… (410)
A. 正态曲线的面积和纵线表 ……………………………………………… (410)
B. t 概率分布表 …………………………………………………………… (420)
C. χ^2 概率分布表 ……………………………………………………… (421)
D. F 概率分布表 …………………………………………………………… (423)

上　篇

理论方法篇

1. 西方现代理论语言学的一般理论特征

西方现代语言研究十分讲究方法论，表现出与传统语言研究[①] 有着明显差异的方法论特征。这些方法论特征一方面与西方哲学传统、西方当今社会的总体科学技术发展背景相关，另一方面与西方现代语言理论及认知科学理论的一般特征相联。为要认识和把握这些方法论特征，有必要首先对50年代在美国兴起的现代语言理论的一般特征加以概述。

西方现代理论语言学的兴起应从1957年Chomsky发表他的《句法结构》(Syntactic Structure）一书算起。尔后，在西方（主要在美国）出现了许多语言学理论。其中影响较大的有生成语法、格语法、概括性短语结构语法（Generalized Phrase Structure Grammar)、中心词短语结构语法（Head-Driven Phrase Structure Grammar)、关系语法（Relational Grammar)、功能语法（Functional Grammar）及词项功能语法（ Lexical Functional Grammar）等。虽说这些语言理论之间存在着很大的差异，有些甚至处于理论观点对抗状态，但它们却表现出一些为传统语言研究所不具备的共同特征。这些特征鲜明地反映在语言理论的认识论基础、理论对象、理论目标、理论性质、理论方法和理论表述等六个方面。

1.1. 西方现代理论语言学的认识论基础

以结构主义语言学为主要代表的西方传统语言学理论的认识论基础是哲学上的经验主义和心理学上的行为主义。经验主义和行为主义的认识论思想集中反映在两个问题上：一是关于人的知识是什么的问题；二是人的知识和能力是怎样获得的或形成的问题。在后一个问题上经验主义和行为主义都否认认识主体在认识中的作用，把知识和能力的获得完全归因于后天经验的结果，而知识和能力获得的过程是一系列（经验）刺激——（主体）反应的过程。在第一个问题上，经验主义和行为主义认为知识内容是对经验刺激的摹写。这种认识论在语言研究问题上表现为以下两个理论主张：第一，在语言知识是什么的问题上，认为语言知识是“外在的”(Externalized)，“语言是一种符号系统”；第二，在语言知识是怎样获得的问题上，认为这套符号系统是通过“训练”而形成的一种习惯结果。以这种认识论为背景的语言研究表现出以下几个特点：

（一）否认人脑的特定属性在认识过程中的作用，认为人脑生下来的时候是白板一块，因而不能在人与动物的区别意义上认识和研究语言这一为人类所独具的种属属性，也就不回答也回答不了为什么只有人才能学会使用语言的问题；

（二）只研究具体语言的语言事实，只研究具体语言符号系统是什么样子，语言事实是什么，如句子是什么样子的问题，而不研究这些语言事实的成因（Etiology)，因而不回答为什么人类语言只能是这样而不会是那样，句子为什么只能这样讲而不能那样讲，只能这样理解而

① 传统语言学一语在此处主要指以美国结构主义语言学为代表的以行为主义为认识论基础的语言学派。

不能那样理解这类问题，从而割断了语言与人脑、语言与心理的天然联系，把心理、认知及人脑的研究排除在语言研究之外。

五十年代初，Chomsky 对美国结构语言学理论的行为主义认识论发起了挑战，提出了“心智主义”（Mentalism）[①] 的认识论思想，在认知科学研究和语言研究领域中引起了一场革命性的变革。[②] 这种心智主义认识论思想便成了美国当代语言研究主流的认识论基础。

心智主义认识论的思想内容主要有二：（一）认为语言是人类所独具的一种种属属性，人之所以会说话是因为人生下来的时候，人脑就呈现为一种特定的物质状态。这种特定的物质状态和结构是人类遗传基因预先规定好了的，它在后天经验（语言环境）的作用下，发育成长而进入一种稳定的物质状态，从而具备了说话的能力，获得了某种具体语言的语言知识。[③] 由于其他动物在其基因的规定下，大脑不具有人脑的遗传属性，不会呈现出与人脑相同的物质状态，因而就没有学会使用语言的可能。一条狗即使放在与人相同的语言经验环境中，即使它受到多么强烈和充足的刺激和训练也不会学会讲出人话来。所以，我们完全有理由把语言最终归因于人脑某种特定的遗传属性。（二）与语言相关的人脑的某种特定的遗传属性决定了人有学会任何一种人类语言的可能。这种可能在后天语言经验的作用下，变成了使用某一具体语言的现实能力。因而语言便是后天经验作用于人脑遗传属性的结果，是先天属性与后天经验相互作用的结果[④]。基于这种心智主义的认识论思想，西方现代语言学理论及其相关学科（如心理学、认知科学等）研究兴趣不只是语言事实本身，不只是对语言事实（如句子、语音等）的描写，而是人脑的遗传属性，[⑤] 是语言共项（Linguistic Universals），[⑥] 是关于什么可成为人类可能语言的限制（constraints），进而从这些限制中找出人脑究竟有着什么样的特殊结构致使人具有学会任何一种语言的可能，而动物却没有。许多现代西方语言学家正是在这种认识论思想的激发下，把语言研究看成是关于人脑研究的一个不可缺少的部分，把语言学看成认知心理学的一个重要部分。

---

① 关于 Chomsky 语言观的认识论基础，有人概括成这里说的“心智主义”，有人概括成“自然主义”（Nativism），有人概括成“新理性主义”（Neo-Rationalism），还有人认为上面说的这几个主义都不准确，而坚持用 Chomskyanism “乔姆斯基主义”。

② 西方语言学界和认知科学学界都使用“革命”一词描述 Chomsky 为人类语言研究和认知研究带来的巨大变化。笔者认为，从人类语言研究的历史角度看，从 50 年代的 Chomsky 起在美国和世界其他一些国家语言学界和认知科学界已发生和正在发生的变化来看，的确是一场科学革命，因为无论是在语言学的认识论基础、理论对象、理论目标上，还是在语言学的理论方法上，都不同于以往任何一种语言研究理论。Chomsky 为人类语言研究开拓了一个崭新的研究领域和方向。

③ 汉语“能力”和“知识”之间似乎有比较明显的差别。在英语中“knowledge”一词有时可以理解成“能力”，所以，讲“语言知识”应理解为“实际使用语言的能力”。由于在许多英语文献中常出现“knowledge of language”的说法，甚行文其他一些地方都表述成“语言知识”。

④ Chomsky 的语言观从没否认过后天经验在人学会使用语言中的作用，明确指出后天经验在语言学习中的“触发作用”（triggering effect）和“定型作用”（shaping effect）。Chomsky 只是针对行为主义的白板说反复指出人脑先天结构属性在使人能够学会说话方面起的决定性作用，并始终认为人之所以会说话是人脑遗传属性和后天经验相互作用的结果。

⑤ 人脑的遗传属性既然是生物学意义上的属性，西方现代理论语言学又称作“语言生物学”（the biology of language）。

⑥ 语言共项指的是人类语言共同遵循的普遍规则和原理。这里应该首先区分两个问题，一是有没有语言共项？一是什么是语言共项？回答第一个问题可以从这样一个问题着手：人类语言有没有限制？如果认为有限制，这些限制说明人类语言必须遵循共同的规则和原理，这就是语言共项。承认了语言共项之后才有什么是语言共项的问题。

1.2. 西方现代理论语言学的理论对象

既然西方现代语言研究者大都接受 Chomsky 心智主义认识论思想，承认人脑固有遗传属性在限定可能人类语言上、在规定语言知识或能力的获得路线和方式上的作用，他们的研究对象不再像结构主义那样停留在对各种具体语言是什么样的描写上，而是通过对各种具体语言，各种语言中的语言事实和现象以及它们之间相互关系的研究，探索人脑的奥秘，揭示人脑究竟有着一些什么属性使人成为一个会讲话的人。西方现代理论语言学及其相关学科的最终研究对象不是语言，而是人脑。用 Chomsky 的话讲得具体一点，他们要回答以下三个称作语言研究中的“柏拉图问题”：

(1) 人类语言知识或能力是什么？
(2) 人类语言知识或能力是怎样获得的？
(3) 人类语言知识或能力又是怎样使用的？

对于第一个问题，西方现代理论语言学不再像结构主义那样满足于回答说“语言是用于交际的符号系统”,① 而认为语言知识或能力是人脑的一种物质状态，是在人脑中构筑起来的一个“计算系统”(computational system)②。对于第二个问题，西方现代理论语言学也不再像结构主义那样回答说语言知识和能力是靠“刺激—反应”训练形成的行为习惯，而认为语言知识和能力的形成是人脑在后天语言经验的作用下，从一种初始状态 (initial state) 进入一种稳定状态 (steady state) 的过程和结果。也就是说，语言知识获得的过程，语言能力形成的过程是人脑在后天经验作用下沿其遗传规定的方向发育成熟的过程③。关于第三个问题，西方现代理论语言学也不再象传统语言学那样认为语言是随意④，而认为是在某些与人脑及整个认知系统固有属性相关的规则或原理的支配下运用的，语言行为也是由规则支配的 (rule-governed)。

① “语言是用于交际的符号系统”是以往语言学著作中极为常见的说法。这样说本身并没有任何不当之处。但是作为科学研究对象的本质定义，至少有三个不妥的地方。第一，说语言是交际工具，这是从语言的功能方面给语言下定义。功能定义法只有描写上的意义，不能从事物的本质上形成界定。这就和说自行车是一种交通工具一样，不能同汽车区别开来一样，因为汽车也是一种交通工具，交通工具不是自行车区别于自行车以外事物的属性。第二，语言不是唯一的用来进行交际的工具。人们除了使用语言进行交际之外，还用体态语言、情感语音等进行交际。第三语言不只用于交际，还用于思考，用于自我心理调节等。

② 把人脑的系统比成是一种“计算系统”大有人在。但是，把语言能力的形成和知识的获得说成是人脑某种特定状态的出现和一种计算系统的建立要首推 Chomsky。

③ 这个人脑发育成熟的结果是像婴儿身体各部位发育成长一样会长出一个可主管说话的“语法”来．西方理论语言学中有“语法生长”(the growth of grammar) 一说。

④ 语言有约定俗成的一面，比如说，汉语把桌子这件东西叫“桌子”而没有叫成别的什么，英语叫“table”而没有叫成别的什么。但语言却有许多无法约定的东西，本文后面要讲许多例子说明这点，这里只举两个例子，似乎足可以说明问题。在语音系统方面，无论如何约定，人类语言中不会含有 [bp] 这样的辅音串，也不会允许有 [我打算我去] 这样的句法结构。传统语言学没有把这类约束看成是人类语言中很重要的甚至是本质的一面。

总之，西方现代语言理论的理论对象完全不同于传统语言理论。前者研究语言事实的成因，研究人脑，研究可能人类语言的限制和共项；后者研究个别语言，研究语言事实自身，研究语言间的个体差异。

1.3. 西方现代理论语言学的理论目标

西方有一种关于科学理论的分类法。这种分类法把科学理论分成两大类。一类科学理论旨在描写科学事实，对科学事实提供精细的分类描写，人们称之为描写性理论（descriptive theory）。门捷列夫的化学元素周期表就属此类，结构主义语言学也属此类。另一类科学理论旨在为科学事实的成因提供理论解释，回答什么原因使事实成为这种样子，而不是那种样子。这种理论叫解释性理论（explanatory theory）。牛顿的物理学说，爱恩斯坦的相对论和孟德尔一莫根的遗传学说都属解释性理论。描写性理论追求“描写上的充分性”（descriptive adequacy），解释性理论追求“解释上的充分性”（explanatory adequacy），在解释充分性的基础上追求描写上的充分性。当然，一种理论通常会同时具有描写的理论性质和解释的理论性质。不过，以描写概括事实的广度为最终目标的理论，即使含有解释性的理论色彩，仍旧是描写性的理论；而以解释事实成因为最终理论目标的理论，即使含有必要的描写性理论色彩，仍旧是解释性理论。本文所说的“理论语言学”与“解释性语言理论”在使用意义上大体一样。

大多数西方现代语言学理论都具有鲜明的解释性理论性质。这表现在以下几个方面。第一，这些理论不拘泥于对个别语言的具体语言事实的描写。不停留在对具体语言的具体语言现象的归纳分类描写上，而是从对具体语言现象的观察出发，研究各语言现象和事实间的内在联系，在联系中寻找事实的系统性成因。第二，解释性语言理论或理论语言学以研究各种语言间的差异为起点，把差异看成某一语言共有属性的变体，在通过对变体的研究，探察变体的本源，寻求语言共项和普遍语法原理（Universal Grammar Principles，简称 UG Principles）。第三，解释性语言理论在语言直觉问题上也与描写性语言理论不同。描写性语言理论凭借研究者语言直觉和研究对象的语言直觉确认语言事实的真伪，认识和描写相关语言事实的特征。而解释性语言理论则不然，它不只凭借研究者和研究对象的语言直觉，更重要的是要揭示这些语言直觉的内容及其形成的原因，回答一个人为什么觉得“张三睡着了”是一句话而“张三睡着了李四”不是一句话这类问题。在对待语言事实和语言直觉的态度上，描写性语言理论与解释性语言理论表现出鲜明的差异。描写性语言理论研究的是人说出来的或听到的句子，解释性语言理论研究的是说或听句子的人。

1.4. 西方现代理论语言学的理论表述特征

任何理论都得取某种方式来表述其理论内容。在社会科学中，多以自然语言语句表述，在自然科学中，除了用自然语言语句表述外还用形式化的手段（如公式、数字、图形等）表述。西方现代理论语言学在理论表述上有着鲜明的自然科学色彩，大多数都追求一种类似数学或

计算机系统的形式化的表述系统，都是典型的形式化理论系统。这种形式化的理论表述主要表现在两个方面。

第一，西方现代理论语言学都是公理性系统。这种公理系统含有一些初始元(primitives)，这些初始元无须也无法给出定义，就像数学中的数字符号“1”和“2”那样。人们认为人类的语言知识和人类的数学知识一样也含有一些原本无须也无法定义的概念。比如说，某些“范畴”(category)概念，某些自然类属(natural class)等。像什么叫“施事”(agent)，什么叫“受事”(patient)，什么叫“谓词“(predicate)，什么叫“是”(is-a)等就可能是自然语言中的初始元。由于语言理论是关于自然语言的描写，因而语言理论的表述是关于自然语言的表述，如果语言理论的表述仍旧使用自然语言，那么结果就会出现用自然语言描写自然语言的情景。为了在表述和表述对象之间“拉开”一个心理距离(psychic-distance)，就有必要创造出一个“元语言”(meta-language)。与其他学科相比，元语言在语言研究中更为重要。在物理学中，理论表述与理论表述的对象之间有着一个天然的界限，表述是语言的，表述对象是非语言的物理世界。在数学中，理论表述用的是数学语言，表述对象是物理世界的形式属性。而在语言研究中，理论表述和表述对象之间就没有这种天然的界限。这套元语言系统又不能随意地创造出来，而必须在人类知识能力范围之内建立，必须具有人类的可读性，因为无论理论做何等表述都得让人看得懂。这样，建立这套元语言系统就不可避免地要依靠人类自然语言系统中已有的或隐含的初始概念。当然，像有的西方语言学家和哲学家说的那样，数学语言、几何语言、物理公式语言等本身就是人类自然语言的一个部分，并非人类语言之外的创造。所以，西方现代理论语言学的公理系统实际上像数学是从人类自然语言中挖掘出一些初始元后建立起一套公理性系统一样，也企图在人类自然语言中挖掘出一些可用于描写语言的初始元而后建立起另一套公理系统。在这方面，西方现代理论语言学广泛地借用了现有公理系统(如数学、几何、形式逻辑等)中的初始元。有了初始元不等于就有了公理系统的全部。还需要在已确定的初始元的基础上规定出一些“算法”(algorithms)，然后在已确定的算法基础上规定出一些“公理”(axioms)来，最后形成完整的公理系统。这个公理系统和所有公理系统一样具有自足性(self-evident)。认识西方现代理论语言学应该从公理系统的角度去认识。但是，包括公理性已经很强的生成语法、概括性短语结构语法和词项功能语法，甚至包括本身就是数学公理系统的“蒙太古语法”(Montague Grammar)以及采取谓词演算形式的形式语义理论(Formal Semantics)在内，都还没有把这个用来描写语言的公理系统最终完整地建立起来，主要原因和困难在于什么应该是所需的初始元。也正是出于这种原因，西方现代理论语言学家们都在努力寻找和确立初始元，而也正是在这个问题上，各家有各家的看法。

西方现代理论语言学理论表述上的第二个特征是理论系统的操作性，就是说他们想要建立起来的语法理论是一个像计算机程序那样的操作系统。它们所要建立的是一个可对数据材料进行处理加工的程序系统。这一点和传统语言理论比较一下，是很好理解的。

2. 西方现代理论语言学理论方法的总体特征

如上节所述，大多数西方现代理论语言学都普遍接受 Chomsky 的心智主义，都认为人脑的某种特定的生物遗传属性决定了人且只有人才能使用语言，语言正是这种人脑生物遗传属性的直接产物和表征，进而把语言研究的对象界定为人脑的“语言器官”(the Language Faculty)，通过研究人类语言为人脑的特定结构建立一个解释性的理论模型。相应之下，在理论方法上表现出以下两个总体特征：

（一）抽象理论模型法

（二）模块理论法

2.1. 抽象理论模型法

为弄清什么是抽象理论模型法，为什么西方现代语言理论大多数都采用这种理论方法，我们不妨先从西方现代理论语言学所界定的理论对象——人脑讲起。

人脑首先是一个实体，它具有一切实体所具有的物理属性，因而人们可以在物理这个层次上去研究它。可是，人脑又是一个生理实体，人们也可以在生理这个层次上去研究它，例如神经生理学和神经语言学就是这样。但是在这个层次很难找到人与动物（包括高等动物）的本质差别，因而在这个层次上研究人脑，无法认识和解释人脑的特定属性，无法把人脑与动物脑从本质上区别开来。所以说，语言作为一种人类种属属性首先不会体现在这两个层次上[①]。以人脑为最终研究对象的西方现代理论语言学便首先不会对人脑的生化物理属性感兴趣，而对与语言直接相关的人脑的“高级”层次——“语言认知层次”感兴趣。作为科学研究对象，生化物理层次上的人脑或人脑的生化物理属性则与语言认知层次上的人脑或人脑的语言认知属性不同。前者可凭借我们的认识工具，通过经验（实验、实证的方法直接观察，如通过对动物、病人及死者器官的解剖等）实证方法观察研究，而后者由于受人类道德及自身认识的限制则不能，另一方面也无法用研究生化物理属性的方法在语言认知层次上认识和研究人脑。这样，人脑作为理论语言学研究对象就成为不可能（至少现在和很久以后的将来）靠经验实证方法来认识的实体。这就涉及到科学方法论上一个带有根本性的问题：人能不能认识无法在经验（包括实验室经验）中直接（包括通过科学仪器）观察到的东西？如果能，又如何去认识？如果有了办法去认识，又在什么情况下，才算是有了认识？对于这些问题，人们实际上已在许多重大的科学理论中作出了回答。在能不能的问题上，回答是肯定的。在如何认识和怎样才算认识的问题上，回答是建立理论模型的方法。

① 从科学研究的认识过程上讲，人们无法首先在生理神经层次上看到语言的实现，这不等于说语言实现的结构层次是这样的顺序。

以人们业已认识了的原子为例。对于称作“原子”的这个实体，人们并没有办法“看见”，无论现代物理实验手段、观察认识工具多么发达，多么丰富，到目前为止（也许直到永远），人们不可能像看见病毒或 DNA 那样看见任何一种原子。那么，人们究竟是怎样发现原子这一物理实体，又是怎样确认它的存在的呢？回答是靠一个关于什么是原子的抽象的理论模型或范式（paradigm）。这个关于什么是原子的理论模型是由许多关于原子属性的数据构成的，而各种数据又都是靠仪表、仪器测定的。比如说，仪表 A 读出数据 a，仪表 B 读出数据 b，仪表 C 读出数据 c 等等。这些从仪表 A、B、C 读出的数据 a、b、c 等就构成了某一原子的模型。假如这个模型表达为 [a，b，c，. . .]，那么当某物 X 作用于仪表 A、B、C，给出数据 a、b、c 后被观察者所读出时，观察者才感知原子的存在，才“看见”了原子，找到了原子。可以说，原子模型的建立先于原子实体的发现，对原子存在的确认先于原子的发现，没有原子模型的建立便无法找到和“看见”原子。在科学史上运用抽象理论模型法得到重大科学发现的一个例子是孟德尔和摩根的遗传学理论。孟德尔当年提出遗传基因论的时候，并没有真的看见基因物质的存在，而是在对豌豆遗传变体性状观察的基础上，抽象出一个模型，而后人们“按图索骥”，根据孟德尔和摩根提出的理论模型，在显微镜下辨认出了那种叫 DNA 的东西。

西方现代理论语言学家们认为他们现在所从事的事业就和孟德尔、摩根当年建立基因模型一样，在用同样的方法为人脑的认知系统建立一个抽象的理论模型。说他们的方法一样，可以从以下几个方面看出。第一，孟德尔当年看不见现在看得见的那个主管豌豆植株属性的称作 DNA 的物质实体，语言学家们看不见主管人类语言属性的那个将来可能看得见的物理实体；第二，孟德尔没有停留在对豌豆隔代植株间的差异和性状的观察描写上，而是要寻找植株变异的原因，从变异中找出豌豆的种属属性来；西方现代理论语言学家们没有停留在对语言间的差异和特性的观察和描写上，而是要在语言间的差异中寻找人类语言的限制，从中认识确定人类语言的本质特征。第三，孟德尔没有也不能直接观察豌豆的基因实体，而是观察研究与基因有关的表象特征，西方现代语言学家没有也不能直接观察到人脑主司语言部分的真实性状，而是观察人脑的表象产物——语言中的句子。第四，孟德尔没有观察研究豌豆的全部属性，而只是在理想化条件下集中在几个相关的属性上，从中找出各属性间的天然联系，建立起一个抽象的理论模型。西方现代理论语言学家们没有观察研究人类所有语言的所有属性、所有人说过的和将要说的所有的句子，而是忽略语言使用者的个体差异，通过对某些与理论相关（theory-related）的语言事实观察研究，从中找出事实间的天然联系，建立一个抽象的理论模型。西方现代理论语言学家们希望能沿着孟德尔研究遗传基因的方法、道路为人脑提供一个同样具有理论意义的理论模型。

讲到这里，我们不由得想起了祖国中医的理论方法。在中医理论中，研究者并没有分析性地把人体打开，通过经验的直接观察的方法解释病理，而是在观察研究体征表象、症状及它们之间的相互关系，用推理的方法建立了一个系统，这个系统原本也是一个抽象的理论模型，只不过是最近人们才按照几千年人们画好了的那个经络图，找到了并看见了那个称作经络的实体。西方现代理论语言学家好像我们的古代医学家一样，在为人脑的认知器官勾画一幅经络图。笔者认为在很多方面西方现代理论语言学和中医理论十分相似。这不能不说是一个很有趣的科学历史现象。

2.2. 模块式的理论模型

西方现代语言理论所要建立的语言理论模型是典型的模块系统（modular system）。任何一个模块系统都是由几个相互独立又相互关联的子系统构成的有机系统。比如说，我们要建立的模块系统是 A ，那么，A 则由它的子系统 a、b、c 构成，a、b、c 在总系统 A 中各自有各自的作用。它们之间有着明显的差别，但不是平行并列的关系，而是相互作用，有机地联系在一起。在 a、b、c 子系统之间存在着"输出—输入"的推导关系。子系统 a 的输出 a′可能是子系统 b 的输入，当子系统 b 对 a′进行加工处理后，可能会向子系统 c 输出 b′，然后子系统 c 对 b′进行加工，输出 c′。而 c′便成了整个系统 A 的最终结果输出。子系统与子系统之间的相关部分叫接口。一个子系统能够和应该输出什么与另一个子系统能够接受和加工什么必须一致，必须满足接口条件（interface conditions）。每一个子系统还可以由更多的子子系统构成，各子子系统之间也呈"输出—输入"的关系，也必须满足接口条件。每个最小的子系统则由一些规则或原理（principles）构成。每个规则或原理除有自身的规定外，都必须有使用的条件，规则与规则之间不是平行并列的叠加，它们之间互相关联互相依存，表现出一种系统性和操作性。

由于子系统 a、b、c 是系统 A 的内部系统，我们称之为 A 的内模块（internal modularity）。建立语言理论模型就得把语言这个系统的内部结构按照内模块思想建立起来，确定语言系统内究竟是由哪些既相互独立又相互关联的子系统组成，各子系统各自的任务和分工是什么，它们之间的接口条件又是什么，每一个子系统又是由哪些规则构成的，各规则间的相互关系是什么，各规则的使用条件又是什么等。

然而，有了一个系统的内模块，并不等于建立了系统的全部。因为我们要为语言建立的这个系统还是整个认知系统（cognitive system）中的一个子系统。西方现代语言理论明确地把语言理论研究对象确定为人脑，即人脑主管包括语言在内的认知系统。这样语言系统又和认知系统呈现为一种模块关系。这种模块关系叫外模块（external modularity）。因而，建立语言系统不得不考虑语言系统在整个认知系统中的作用和意义，不得不弄清人的整个认知系统的模块性，就是说，认知系统作为一个总系统由哪些内模块子系统构成，各子系统间的接口及其接口条件是什么。

西方现代语言学家的全部工作可以说都是在按着这种语言系统的内模块和外模块的思想为人类语言建立一个抽象的模型。这势必与传统语言研究在总体方法论上表现出不同。抛开西方现代语言理论与传统语言理论在研究对象上的差异不谈，从它们所建立理论的系统性上可看出这一明显的差别。传统语言理论，由于隔断了语言研究与心理研究及其他认知系统研究间的联系，没有把语言系统看成是整个认知系统的一个有机部分进而把语言放在整个认知系统中进行研究，因而按照这种思考所建立起来的语言理论或语法与认知系统及这个认知系统中的其他子系统之间没有外模块关系。它的作用、它的存在意义不受认知系统中其他子系统的约束和规定。所以说，语言理论或语法是什么样子及应建成什么样子，不必放在人的整个认知系统中考虑，不必在整个认知系统中从同其他认知子系统的相互关系上考虑。这是其一。其二，在传统语言理论的语法体系中，的确有许许多多规则，但这些规则大多数是针对

某些语言事实的共同特征建立起来的。它们之间的关系是平行并列的，没有相互依存的关系，也没有使用上的条件，整个语法系统没有系统的总体任务，规则可以是游离于系统之外的孤立的规则，规则中所使用的概念可以是系统之外的孤立的概念。一个规则或一组规则的建立往往只对某些事实负责，而忽视或不承认一个语言事实或一组语言事实同一个语言事实或另一组语言事实间的依存关系。因而规则也就没有使用的顺序，而规则自身也就没有操作性，不会对任何东西进行加工处理，而只是静态地陈述和描写，规则与规则之间、子系统与子系统之间也就没有“输出一输入”的推导演绎关系。

那么，西方现代语言理论究竟是怎样按照内模块和外模块的理论思想为人类语言建立抽象理论模型的呢?它们迄今为止所建立起来的理论模型的内模块和外模块究竟是什么样子呢?

2.3. 语言理论模型的外模块系统

语言系统的内模块及外模块显然都不能通过经验或实证的方法来辨认和确定，而只能通过对这个系统的总输出，即语言事实、语言现象、语言事件、语言运用行为的研究来确立。这自然便引出了下面两个问题：

(1) 我们赖以建立语言理论模型的语言事实、语言事件究竟是什么?

(2) 如何依据语言事实来建立抽象的语言理论模型?

让我们先讨论第一个问题。任何科学理论都是关于所界定研究范围或对象所及事实的理论。自然科学是这样，社会科学亦如此。但与语言理论相关的科学事实却不同于自然科学中的事实，也不同于社会科学中的事实。这是因为，语言事实一方面有其物质的属性，如语言的声音，文字符号等，另一方面语言还有其意识、精神的属性，这就是它的意义。由于语言所表现出的这种双重属性，语言事实是什么，语言事实在哪里去寻找，本身就是一个重大的理论问题。以语言中最根本的事实或事件单位句子为例，对什么是句子，句子在哪里的问题，不同的理解与回答会引起差别很大甚至性质根本不同的理论。也正是在这个问题上，西方现代语言理论首先表现出与传统语言理论明显的不同。请看下面这句汉语：

(3) 张三准备去。

任何一个会说汉语的人都会感觉到(3)是汉语中的一个句子。但是，说(3)是一句话或一个语言事件究竟是什么意思，又在什么情况下它才成为一句话、一个语言事件的呢?这在传统语言研究中不曾被认真讨论的问题，在现代语言研究中却成了必须回答的问题。(3)只有在以下两个条件下才会成为一句汉语、一个汉语事件：第一，这串声音或文字符号必须由讲话者或听话者讲出来、写下来或听话者辨认出来；第二，讲话者或听话者同时“懂得”这串符号所承载的意义。前者是发生在讲(听)话者的发音器官及听觉器官的物理感知事件；后

者是发生在讲（听）话者大脑中的思维意识活动。只有当这两个事件共时发生时[①]，(3) 才会成为一个句子，才是一个完整的语言事件或事实。所以，观察研究语言事实就等于观察研究人是怎样发出和辨认声音符号及从所发出和辨认出的符号中传达和听懂了什么。所以说，语言事实只能是感知的和意识的共时事件，是牵涉到认知系统多个子系统的认知事件，不会是单一的发音器官的物理事件，也不是单纯的意识事件。既然语言事件的发生包含一个物理声音形态的辨认过程，也包含着一个意义辨认和理解的意识过程，那么这两个既区别又联系的过程肯定是由既区别又联系的两个认知子系统来分别实现。基于这种总体考虑，西方现代语言理论把探索这两个认知子系统的界限及相互关系看成建立语言理论模型必须首先回答的问题。把语言看成既有物理声音的一面又有意识思维的一面并非西方现代语言理论的新发现，传统语言理论早就有之，甚至是几千年的常识。但是，把句子、语言事实看成是一个物理感知的和意识的共时事件，并把两者统一归因于一个完整的认知系统，从而通过对语言事件的物理属性和过程及语言事件的意识属性和过程以及两者的共时联系的研究，推断出人们不可经验实证认识的那个认知系统的结构和属性，这确实是西方现代语言理论有别于传统语言理论的地方。

然而，把人的认知系统说成是由一个主管发（辨）音的子系统和一个主管意义或意识的子系统构成的仍旧十分笼统。五十年代末，Chomsky 首先发现了 (4) (5) (6) 之间差别的重要理论意义[②]，提出了一个关于人类认知系统的内模块结构及语言系统在认知系统中的外模块关系的重要看法，为后来西方现代语言学家所普遍接受，成为确定语言系统的外模块结构的公设依据，是认识西方现代语言理论的一个重要出发点。

(4) Little white sheep sleep quietly.
　　小白羊安详地睡觉。
(5) Colorless green ideas sleep furiously.
　　无色的绿思想疯狂地睡觉。
(6) Ideas green furiously colorless sleep.
　　思想绿疯狂地无色的睡觉。

比较 (4) (5) (6)，我们会发现 (4) 及其汉译是一句非常自然的英文句子和汉语句子，可任何人在任何情况下都不会说出 (6) 这样的句子。换句话说 (6) 根本就不是语言中的句子。更引起 Chomsky 理论兴趣的是 (5)。它的特点有二：(一) 与 (4) 相比，(5) 在句子结构上和 (4) 没有任何差别，和 (4) 一样合乎英文语法；(二) 但在所表达的意思上，(5) 却无法按常识理解。而 (6) 在结构上是不合乎英文语法的。这样 (4) 和 (5) 为一方和 (6) 自己为另一方，两者之间 形成了一种对照 (contrast)。从这种对照中，Chomsky 得出了这样几

① 讲语言活动中的物理生理感知事件与思维意识事件共时发生，在西方现代语言学中并未见有这方面的明确阐述。笔者在这里提出这点是想从一般常识意义出发，说明语言事件离不开声音形式和意义。不过，这里就会引出另外一个问题——人在无声思维时语言事件是不是变成了单一的思维意识事件。对此，西方现代语言理论没有统一明确的回答。不过笔者认为即或人在进行无声的语言活动时，也要靠主司意义的认知系统和主司发音的认知系统同时参与来实现。主要的不在有声语言活动和无声语言活动之间的差别，而在人的认知系统是否存在着相互独立有相互联系的主司声音的子系统和主司意义的子系统。

② (5) (6) 为 Chomsky 原话，(4) 为笔者所造。

个结论：(一) 人们都有一种独立于意义之外的关于某种语言句法结构的语言直觉，称作语法或句法独立性或自治性 (the independence or autonomy of syntax)；(二) 能够传达意义和能被理解的合乎常识的句子必须合乎语法，而合乎语法的句子不一定合乎常识；(三) 合乎语法的句子多于既合乎语法又合乎常识的句子；(四) 不合乎语法的句子肯定不会传达任何意义。这些分析正说明在我们的认知系统中应该有一个独立于常识意义之外的语法或句法子系统。

在Chomsky这种关于语法或句法自治性的思想启发下，西方现代语言学家反复研究了大量的类似于 (5) 的句子，坚信人的认知系统中至少包含一个称作语法的子系统和一个称作"信念"的子系统。再观察下面两个汉语句子：

(7) 大象吃香蕉。
(8) 香蕉吃大象。

从汉语语法结构上看这两句话没有任何差别，都是合乎汉语语法的句子。差别只在于前一句既合乎汉语语法又合乎我们所了解的常识；而后者只合乎汉语语法却不合乎我们目前所了解的常识. 让我们再看下面 (9—11) 中的汉语例子：

(9) 地球围着太阳转。
(10) 太阳围着地球转。
(11) 太阳转围地球着。

显然，(9) 既合乎汉语语法又合乎我们现在了解到的常识，而 (10) 合乎汉语语法，虽说不符合我们现在所了解到的常识却符合哥白尼之前人们所了解的常识。从哥白尼到现在太阳与地球的相互关系没发生任何变化，变化了的是人们的常识信念系统的内容，从原来以为太阳围着地球转变为地球围着太阳转。不管信念系统如何变化都不会影响语言中句子是否合乎语法，(10) 在哥白尼之前和在哥白尼之后都是合乎汉语语法的句子，(9) 在哥白尼之前和哥白尼之后都合乎汉语语法。而在任何情况下，(11) 也不会成为合乎汉语语法的句子。同理，至于说是"大象吃香蕉"还是"香蕉吃大象"也取决于人的信念系统，当人们以为大象吃香蕉的时候"大象吃香蕉"便是既合乎语法又合乎信念常识的句子，而"香蕉吃大象"只合乎语法不合乎信念常识，当人们以为 (如在梦魇中，在因于现实世界脱节而精神失常的病人口中)[①] 香蕉吃大象的时候，"香蕉吃大象"这句话也就成了既合乎语法又合乎信念常识的句子了。

基于上面的观察和考虑，西方现代语言理论通常把 (4)、(7)、(9) 这类既合乎语法又合

---

① 西方现代语言学还用加上"I believe/think..."及"I don't believe/think..."的方法证明信念系统在语言中的作用。比如说，下面这些汉语句子就很生动的说明了这个问题：

(1) 我不认为香蕉吃大象。

(2) 我认为大象吃香蕉。

把这两句话前面的"我不认为""我认为"去掉就是正文中所讨论的句子。反过来讲，任何一句话的前面都可以加上"我认为"而不会影响句子原来的意义。

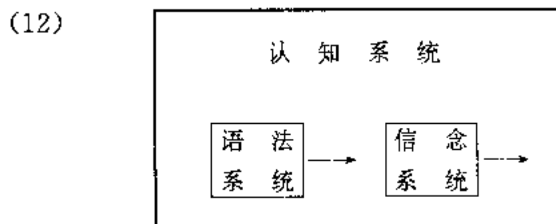
乎常识信念的句子同（5）、（8）、（10）这类合乎语法但不合乎现实世界常识信念的句子放在一起，统称为“合乎语法的”（grammatical）句子；而另一方面，把（6）、（11）这类不合乎语法的句子称作“不合乎语法的”（ungrammatical）句子或“非句子”（non-sentences）。在表述上，合乎语法的句子用正常方法书写，不合乎语法的句子在前面加上一个星号（*）表示成下面的样子：

（6）* Colorless sleep green furiously ideas.

（11）* 太阳转围地球着。

合乎语法的句子又称“可能句”（possible sentences），既合乎语法又符合常识信念的句子亦称“现实句”（realistic sentences）。在可能句与不合乎语法的非句子之间有明显的界限，这个界限使我们认定了认知系统中语法子系统与信念系统的相对独立性。按照这种考虑，西方现代语言学家构想出了下面一个关于人的认知系统的总体框图：

（12）



在这个框架中语法系统与信念系统之间存在着“输出—输入”的关系。语法系统输出所有的可能句，作为信念系统的输入。信念系统对输入的可能句加工处理，输出人们实际运用的现实句。由于可能句总是多于现实句，信念系统对可能句的加工处理实际上是从可能句中按信念常识筛选出不合乎常识信念的句子，最后输出现实句，这些现实句便成为认知系统的终端输出。例如语法系统会输出（13—17）中所有的可能句：

（13）大象吃香蕉。

（14）香蕉不会吃大象。

（15）香蕉吃大象。

（16）a. 地球围着太阳转。

b. 太阳不围着地球转。

（17）太阳围着地球转。

信念系统接受这些句子后，便会选出（13）、（14）、（15）、（16）为现实句输出，而这些现实句也正是我们常常说出或听到的汉语句子。在某种情况下，信念系统也可能把（13）到（17）的句子全都选为现实的句子。西方现代语言理论中常用“现实世界语义”（real-world semantics）和“可能世界语义”（possible world semantics）分别描写这两类句子。（15）和（17）就属于在可能世界中成真的句子，尽管它们在现实世界中不成真。

语法系统的自治性一方面相对于信念系统而言，另一方面相对于语言的运用，即语用

(pragmatics) 而言。语用除了涉及到常识信念系统因素之外还涉及到语境、交际场景、话语篇章 (discourse)、意图、社会文化等方面的问题。如果把这些因素统称为语用因素，它们显然同语言的语法自身固有的属性有着天然的界限。一句话是否合乎语法，不受语用因素决定。像“我不喜欢苹果”这句话即使不能在作为回答“你喜不喜欢香蕉?”的语境中使用，也不会因此失去其合乎汉语语法的性质。至于在什么场合下能说或不能说这句话，取决于语用因素。显然在语法与语用之间也存在着一个界限。这样，在西方现代语言理论中，又在认知系统中圈定出一个称作“语用系统”(pragmatic system) 的子系统。语用系统和信念系统关系十分紧密，两者之间的界限人们尚没有明确的比较统一的看法，这两个系统内的结构尚不能用形式化的方法清晰地刻画出来，所以人们也就常常把语用系统和信念系统混在一起，称之为“信念语用系统”或“语用系统”等。

在 (12) 这个关于认知系统示意图中，我们权且使用了“输出句子”、“输入句子”这样的术语。这在西方现代语言理论中也是常见的表述。但是，把语法系统的输出说成是“句子”的这种表述很不严密，也不是西方现代语言理论中的原有意思。那么，语法系统究竟输出什么呢? 让我们以 (18) 汉语为例：

(18) 鸡不吃了。

这句话和任何一句汉语一样，既有其声音或文字形态一面，又有其意义一面。如果把 (18) 当成一个句子在语法系统中输出，它既表达声音又表达意义。但是，这里就出现了一个问题——这句话是一个歧义句，它的意义可分别描写成 (19) 和 (20) 的样子：

(19) 鸡，(我们) 不吃了。[①]
(20) 鸡不吃 (米) 了。

如果说语法系统输出的是 (18) 这样的句子，得到的只能是一句意义不确定的句子，而使用中的句子在意义上必须是确定的，意义不确定就等于没有意义，没有意义的句子实际上不是句子。只有把 (18) 看成是一种声音或文字形象的描写，同时选定 (19) 或 (20) 作为要传达意义的描写，才能得到一个既有声音 (文字) 形象又有意义的句子。这样，与 (18) 对应的应是 (21) 这样一对包含一个声音 (文字) 形态描写和一个意义描写的表述，或者是 (22) 这样一组包含一个声音 (文字) 形态描写和一个意义描写的表述：

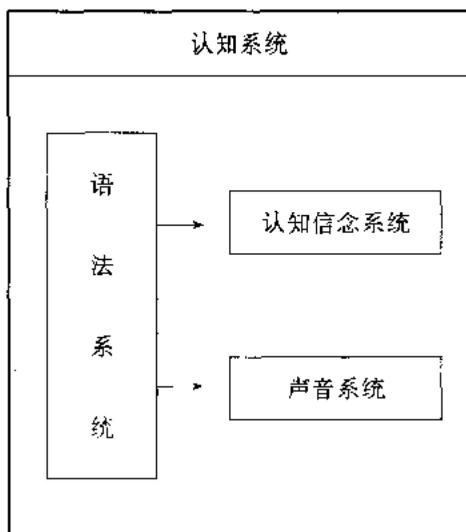
(21) a. [ji bu chi le][②]
　　b. 鸡，(我们) 不吃了
(22) a. [ji bu chi le]
　　b. 鸡不吃 (米) 了

① 这种描写还不能算作是形式化的。

② 此处的声音描写式按西方现代语言理论的要求应该至少用音位学中的“区别性特征”来表达，这里借用汉语拼音表示。

所以，语法系统的输出必须包含两个部分：关于声音形态的描写及关于意义的描写，称作“结构描写”（structural description，简称 SD，为汉语行文方便，以下称“形式化描写”）。然而，只有意义，即语义上的“形式化描写”才会同信念语用系统相关，而声音或关于声音形态的描写与信念语用系统无关。因而，也只有语法中的有关意义的部分与信念语用系统有输出一输入的关系，而有关声音的部分与信念语用系统无关，应和另外一个认知系统中的子系统——声音系统相关联。这就意味着，语法系统内应该有两个输出：一是关于声音形态的形式化描写和一个关于意义的形式化描写。语法系统就应有相应与这两种形式化描写的两个接口，另外认知系统中还应包含一个主司发音辨音的声音系统。那么，（12）那个认知系统结构就应改成（23）的样子。

（23）



西方现代语言学家基本上是在这种认知系统框架中进行语言研究的。但是，由于对于各子系统中的内部模块及各模块中的规则应是什么的看法不同便引起了不同的具体语言理论。生成语法有生成语法的看法，概括性短语结构语法有自己的看法，关系语法有关系语法的看法。甚至同一种语法理论在不同发展时期也有不同的看法。生成语法“句法结构理论”时期的看法就和“扩充式标准理论”时期的看法就不一样，“扩充式标准理论”时期的看法和“管约理论”时期的看法也不相同。但这些理论在语言的语法外模块系统关系上没有本质的差别，尽管在所用术语上各有不同，都普遍接受（23）这个系统框架。语法系统向信念系统输出的是关于句子意义的结构描写，语法系统向声音系统输出的是关于句子声音的结构描写。

2.4. 几个需要说明的问题

受理论语法自身理论目标的规定，西方现代理论语言学不是关于人脑语言器官的现实性描写，一不是关于讲话者生成句子的实际过程，二不是关于听话者理解句子的实际过程，因

而除非特别说明（如少数旨在明确建立“现实性语法”(realistic grammar) 的理论[①]）对生成句子还是理解句子也就不加以区分。不管是生成语法中的树性结构图，还是概括性短语结构语法的“元规则”(Metarules) 和“直接支配规则”(immediate dominance rules)，还是词项功能语法中的“X--阶标论”(X'-Theory) 和“功能结构联结原则”(Principles of Function-Structure Association) 都不能从语言行为的实际过程方面理解和批评。理论语言学是形式化模型理论，应该把形式语法 (formal grammar) 与真实语法区别开来。

另外，西方现代语言学理论严格区分语言能力 (competence)[②] 与语言使用 (performance)。而理论语言学或解释性语言学首先是关于人类语言能力的科学理论，但这并不意味着理论语言学排斥语言使用的研究，而是在现有的理论发展水平上还不见有在理论深度和广度上能同语言能力理论相媲美的语用理论，也不见有在形式化程度上能同语言能力理论相媲美的语用理论。所以，认识西方现代语言理论，有必要在能力与使用或行为的区别意义上认识，而不能用语言能力的理论来评价语言使用的语言理论，也不能用语言使用的理论来评价语言能力的理论；不能用非形式化的语言理论去要求、评价形式化的理论，也不能用形式化的语言理论去要求、评价非形式化的理论；不能用解释性的理论去要求、评价描写性的理论，也不能用描写性的理论去要求、评价解释性的理论；尤其不能用以教授第二语言（外语）为理论目标的教学语法或旨在规范语言使用的规定性语法去要求、评价旨在解释母语能力的理论语言学。

再者，由于哲学、语言学历史背景的原因，西方现代语言学有关于普遍语法[③] 的根深蒂固的认识，十分明确地区分普遍语法和个别语法，所以在认识和评价这些语言理论时，尤其在认识它们的理论方法时，不能按描写性的个别语法去理解、认识解释性的普遍语法。

只要在把握了这些重要的理论性区别后，才能联系西方现代语言理论的一般特征，准确地认识它们的方法论特征。世界上没有统一的、单一的语言学，只有目标各异，性质不同，研究范围有别的各种具体的语言理论，就和没有一个统一的、单一的物理学而只有机械物理、原子物理、电子物理，只有爱恩斯坦的相对论，牛顿的古典物理等理论一样。客观世界并没有给我们分出个什么物理学、化学、数学、心理学、语言学来，也没有在语言学里给我们分出个什么语法学、语用学、语义学、音位学、社会语言学、心理语言学、计算语言学来。这些学科的分界可以说是人为规定的，如何分界并无客观的标准，况且世界的统一性使所有的学科都有间接的或直接的联系，人们只能在约定的同一个研究对象范围内，用约定的统一的认识工具探讨同一个问题。

---

① 见 J. Bresnan："A Realistic Transformational Grammar"。大多数词项功能语法具有比较明显的“真实语法”的性质。

② 关于能力与使用（另有译成“运用”和“行为”的）的区别是西方现代语言理论的一个极为重要的概念，也是区别传统语言理论的一个基本标志。关于什么是语言能力，什么是语言使用乔姆斯基反复做过多次阐述，阐述比较集中的见 Chomsky 的《关于语言问题的反思》(Reflections on Language)、《规则与表述》(Rules and Representations) 和《语言与自然》(Language and Nature) 等。

③ 普遍语法是西方语言理论讨论的“永恒”主题。普遍语法一语最先使用的是法国的“Port-Royal”学派和叶斯帕森。不过，Chomsky 等人赋予了这个概念一些新意。与普遍语法意义相同的说法有“语言获得机制”(Language Acquisition Device)、“人脑初始状态”(the initial state of the mind)，“先验结构”(innate structure) 等。笔者认为最能表达什么是普遍语法的说法是：普遍语法是使人能够在后天经验环境中学会使用语言的人脑遗传属性。

3. 语法系统的内模块结构

3.1. 语法系统的总体任务

由于语言研究的理论目标的规定，西方现代语言理论都不是关于具体语言（如英语、汉语）的具体理论，他们所要建立的和已建立起来的语法也不是关于具体语言的具体语法，而是关于人的语言知识能力是什么，人的语法知识能力是怎样获得的及如何运用的普遍语法，是旨在寻找关于所有可能语言中所有可能句限制的规则系统。这就规定了这种语法系统的总输出或总体任务应该是（24）和（25）：

（24）输出关于所有人类可能语言中所有可能句的“形式化描写”（SD）。
（25）永不生成不合乎语法的句子，或者说，所有的“形式化描写”都应是关于合乎语法的可能句的形式化描写。

举个例子说明。（26）中的各句都是合乎汉语语法的句子，（27）中的各句都是合乎英语语法的句子；而（28）中各句都不是合乎汉语语法的句子，（29）中的都不是合乎英语语法的句子：

（26）A. 张三喜欢他。
B. 张三睡得很香。
C. 张三修车的钳子很好使。

（27）A. John likes him.
B. John slept well.
C. The pliers with which John fixed the car works fine.

（28）A. *张三喜欢张三。
B. *张三睡李四得很香。
C. *张三修车的用钳子很好使。

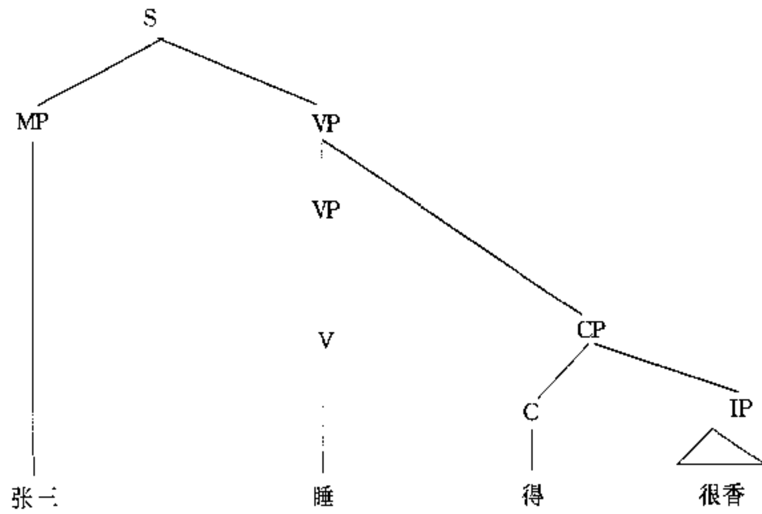
（29）A. *John likes John.
B. *John slept Bill well.
C. *The pliers which John fixed the car works fine.

假如，我们用（30）、（31）、（32）和（33）的比较直观性的形式化描写对（26）、（27）、

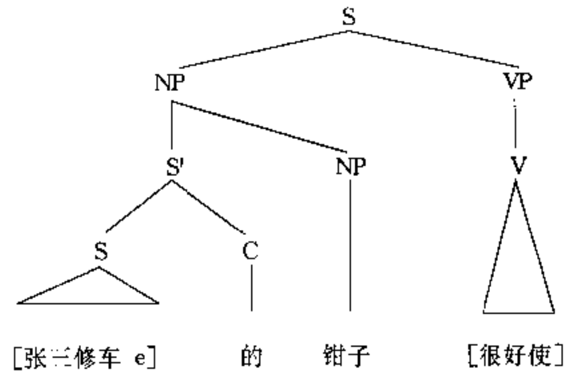
(28) 和 (29) 的意义(忽略语音不计)进行表述[①],其中不带星号的为:

(30) A. 张三$_i$ 喜欢他$_j$

B.

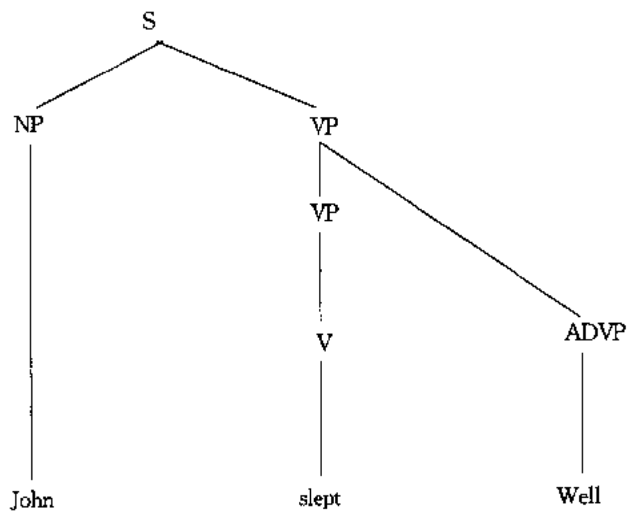


C.



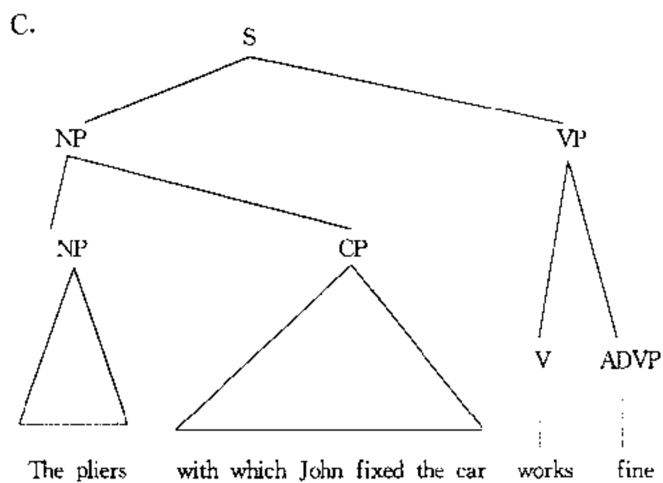
(31) A. John$_i$ likes him$_j$

B.



① 树型结构中的符号及符号间的结构关系不必仔细追究。

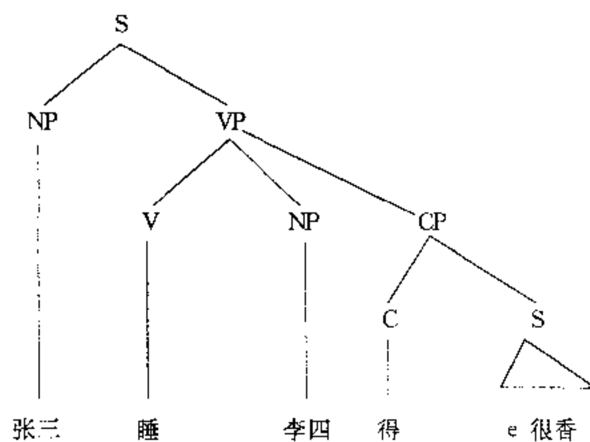
C.



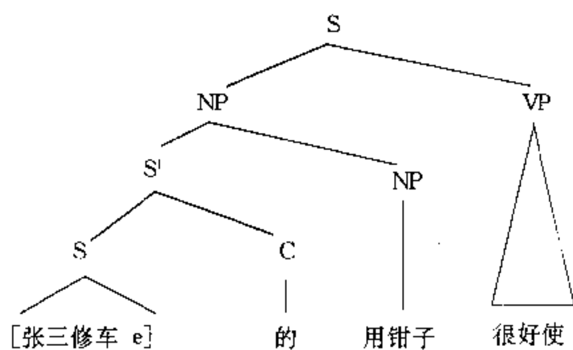
带星号的非句子为：

(32) A. *张三$_i$喜欢张三$_i$.

B. *

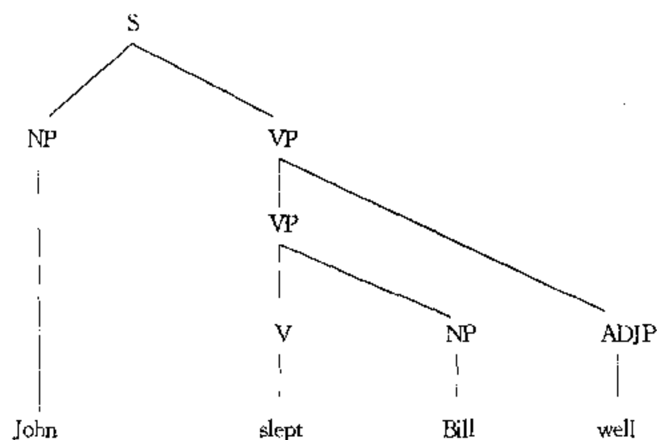


C. *

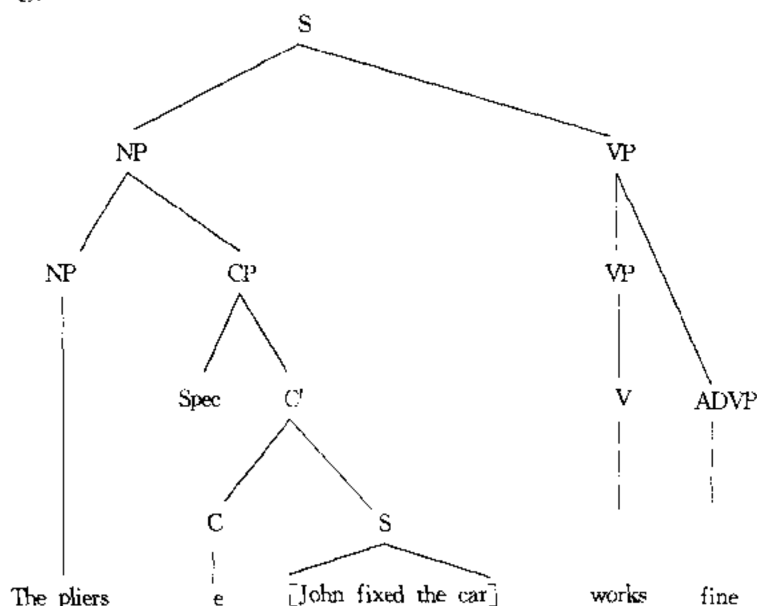


(33) A. * $John_i$ likes $John_i$

B. *



C. *



结构描写式（30A）中的 i 和 j 分别代表两个不同的所指，表示“张三喜欢他”这句话中“张三”和“他”不能同指一个人的意思。(31A)、(32A) 及（33A）所表示的意义也是这样。结构描写式 (30D) 中，各种大写英文字母及其在整个结构树中的位置和相互关系表示各句子成分间的意义关系，比如“张三”和“睡得很香”之间的“主谓关系”由 NP 和 VP 相应于 S 的关系表示。描写式（30C)、(31B)、(31C)、(32B)、(32C)、(33B) 和（33C）大体也表达这种意义。

这些结构描写式是从形式化的语法系统中推导出来的，所有形式、符号、概念都是语法形式化系统中的形式、符号和概念，离开整个语法形式化系统便没有任何意义。这些形式化描写式实际上是语法系统规则推导的结果。语法系统的总输出应该是像（30)、(31）这样一

些关于意义的形式化描写式。这些描写式便成为信念系统输入①，但这个运算系统不会生成像(32)和(33)这样一些对应与不合乎语法的非句子的形式化描写式。在大多数西方现代语言理论著述中人们常常遇见“应生成合乎语法的句子”和“不应生成不合乎语法的句子”等类似的说法，其中“合乎语法的句子”和“不合乎语法的句子”一语，应分别理解为“关于合乎语法的句子的结构描写式”和“关于不合乎语法的句子的结构描写式”。但为了行文的方便，我们还是像大多数西方文献那样把系统生成的形式化描写式简称为“句子”。

3.2. 语法系统的内模块系统形式

语法系统的内模块结构、各结构内的规则、规则使用条件、规则使用的顺序都应为这个语法系统的总任务服务。就是说，语法系统内应由哪些部分构成，各部分之间的关系，各部分又有哪些规则，各规则之间的关系和条件都应从如何实现语法系统的总任务来安排设计。西方现代语言学家按照这样的理论考虑在近四十多年的时间里，提出了不少关于语法系统内模块结构的看法，其中在西方影响最大的要算生成语法的“管约论模型”(Government and Binding Theory Model)以及最近Chomsky提出的“最简语言理论模型”(A Minimalist Program for Linguistic Theory)。此外，有60年代的“生成语义学语法”(Generative Semantics)，60年代到现在的“概括性短语结构语法”，70年代的“格语法”，80年代的“关系语法”(Relational Grammar)，80年代至今的“词项功能语法”等。这些语言理论，在语法系统中的内模块究竟有哪些，它们之间的相互关系是什么，规则有哪些，它们的相互关系是什么，各自使用条件又是什么等问题上有着很大的差异。但是，所有这些理论模型语法系统内模块的构建上都无一例外地遵循以下几个原则。

第一，所有的规则都是形式化的规则。这个要求集中表现在以下几点：(一)规则全部是用“形式构件”(formal objects)表述构成。这些形式构件和数学中的数字符号、运算符号，几何学中的点、线、面、体、角，符号逻辑中的逻辑符号、公式，计算机科学中的数据、运算等自然科学中的形式化表述具有同等意义。由于各家的看法不同，形式构件的选择也不完全相同。大体上分两大类：(A)数据/表征类和(B)运算/操作/推演类。数据/表征类多见以下几种：

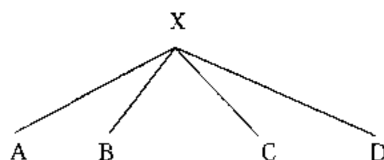
(1)标识符(label亦称节结(node))。如N、NP、V、VP、IP、e、t、John、likes、PRO、the、to等。

(2)标识字符。如[agent]、[theme]、[location]、[goal]等。

(3)关系符号。如下式中的各标识符之间所处的各种结构关系：

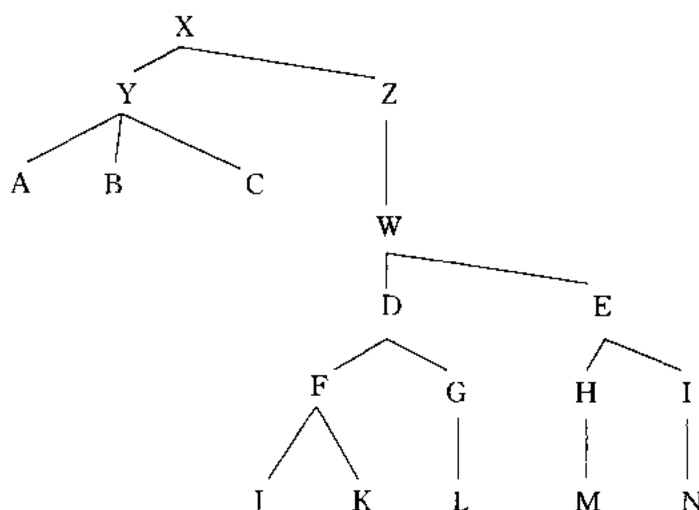
① 语法系统输出的一个描写式对于信念系统来说好像是一条命令语句，让信念系统按照这条命令指示(instruction)去理解句子的意义及句子与现实世界的关系。

(A) 线性关系 (linear relation):



其中，A、B、C、D 各标识符呈线性关系。对于 A 与 B、C、D 之间的关系有的人用标识字符“先于”(precedence) 表述，说成“A 先于 B、C、D”。

(B) 管辖关系 (domination):

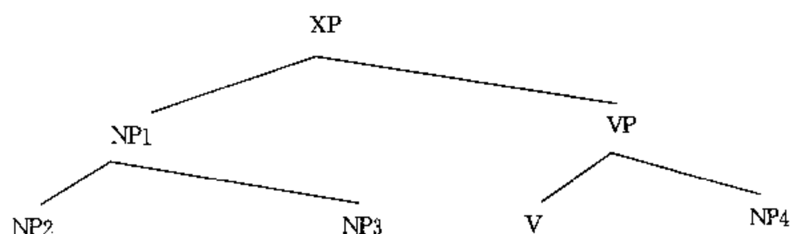


其中，X 管辖 (dominates) 其下所有的标识符。X“直接管辖” (immediately dominates) Y 和 Z；Y 直接管辖 A、B、C；Z 直接管辖 W；F 直接管辖 K 和 J。X 不直接管辖 A、W、L 等，以此类推。根据上述标识符、标识字符及关系符号可以推导出更复杂的数据/表征类描写式来。比如，(C) 和 (B) 就是其中常见的两种。在下面的定义条件下，NP1 统领 VP、V 和 NP4，而 NP2 和 NP3 都不统领这些结构成分：

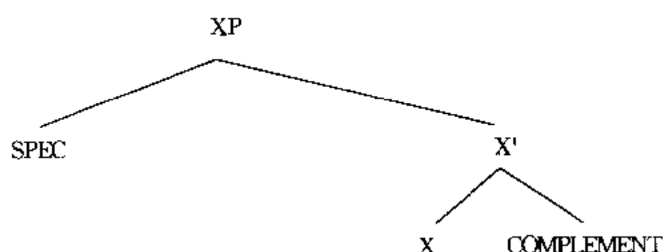
统领条件：X 统领 Y，仅当且当：管辖 X 的第一个短语标记符管辖 Y.

我们可以看出，在 (C) 中，管辖 NP1 的第一个短语标识符是结构中最顶端的那个 XP，而这个标识符又管辖 VP、V 和 NP4。因此，按照上面的规定，我们可以得出 NP1 统领 VP、V 和 NP4。而管辖 NP2 的第一个短语标识符是 NP1，NP1 不管辖 VP、V 及 NP4。因此，NP2 不统领 VP、V 及 NP4。同理，NP3 也不统领 VP、V 和 NP4。这个成分统领的概念为美国语言学家 Ross 60 年代所创，在西方语言理论中一直被广泛用在建构语法系统中。统领关系一方面具有很强的事实描写力，另一方面有着很强的形式化描写力和理论解释力。

(C) 统领关系 (c-command):

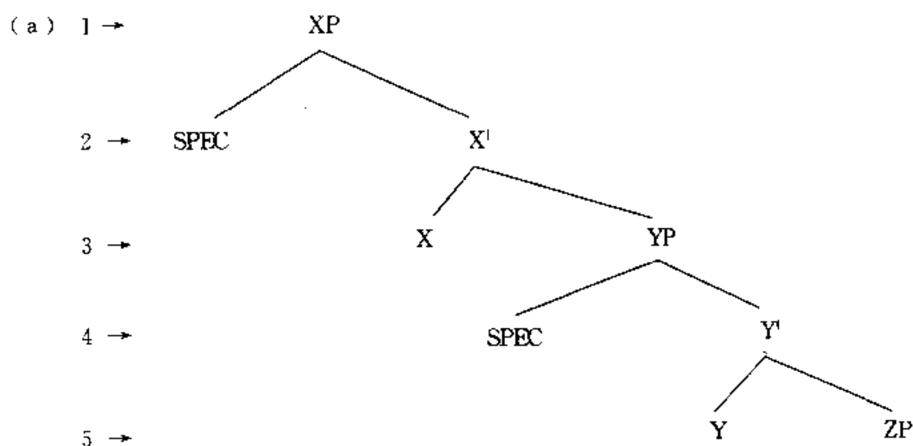


(D) X一阶标关系结构



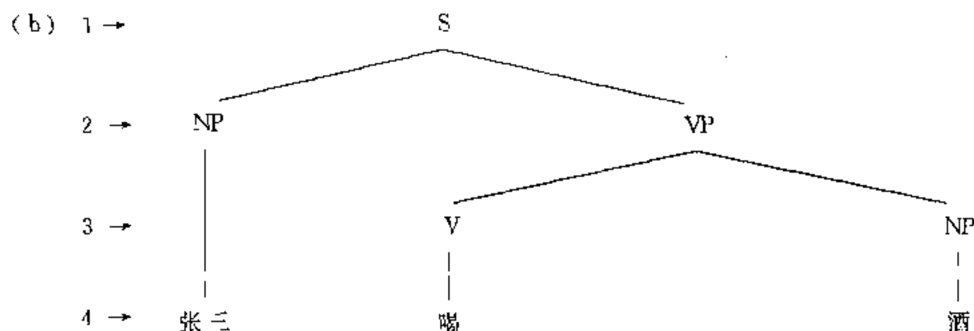
在这个结构形式描写式中，X 是一个标识符，称作“中心词”(head)，COMPLEMENT 也是一个标识符，两者呈线性毗邻关系 (adjacency)，X' 也是一个标识符，与另一个标识符 SPEC 呈线性毗邻关系。XP 是一个特殊的标识符，专门用来标识短语结构，称作“最大映射”(maximal projection)，就是说它是中心词标识符的最大结构短语扩展。这种 X一阶标关系结构在生成语法、概括性短语结构语法、词项功能语法等理论中得到极其广泛的应用，最早出现在 Chomsky 和 Jackendoff 等人的著述中，有着极强的理论解释力。

“层级关系”(hierarchy) 在形式化描写式中也是一个很常见的结构关系。这种关系与线性关系对应，表述不同平面成分之间的关系，也用来表述某些语法功能上的优先性(priority)。在下面 (a) 这个结构描写式中，就含有五个结构层级：



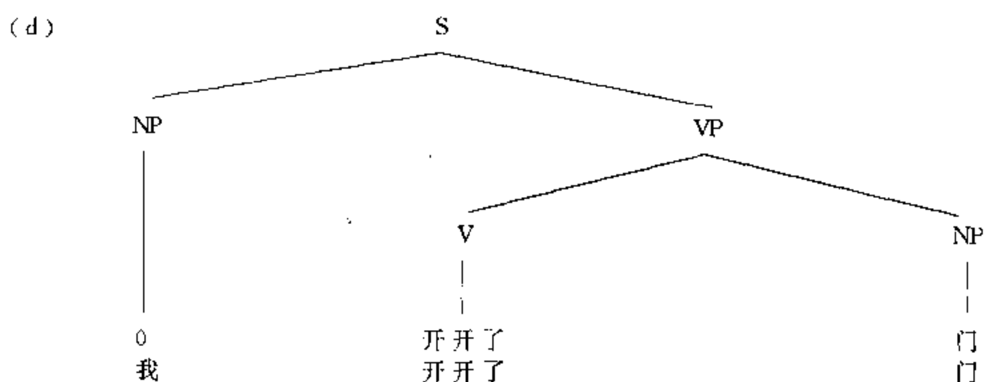
依据层级关系可界定出所需要的、而线性关系所不能表述的东西。比如说，(b) 中所表示的是在所有语言中都可见到的“主- -宾”不对称现象 (asymmetry)：

这样，人们关于主语“张三”与动词“喝”的关系比宾语“酒”与动词“喝”与的关系更为紧密的语感就在层次 2—3 上得到了形式化的刻划，而在层次 4 上或层次 1 上则不能。下面是另一种表达层级关系的方法，人们常用来描写某些标记意义的标识符要比另一些标记意义的标识符在进入规则使用时有优先权。



(c) 施事 (agent) >受事 (patient) >目标 (goal)

比如说，在 (d) 这个结构描写式中有一个结构位置 0 尚未插入任何词项，任何一个放在这个位置上的词项必定充当“施事者”(agent) 的角色，因为，“施事者”应优先于“受事者”得到满足，而不会是“目标”：



标识符也可用来标记语义方面的特征。比如说，在西方现代语言理论中，几乎所有的语法系统都使用下面一些标识符描写语义特征：

[+wh]：表示疑问特征

[—wh]：表示非疑问特征

[+animate]：表示有生命事物的特征

[—animate]：表示非生命物特征

[+pronominal]：表示指代性特征

[－pronominal]：表示非指代性特征

[＋male]：表示阳性特征

[－male]：表示阴性特征

这类特征标记法多采用二元制特征标记法，即给定一个特征标识符，只有两个对立的值，或正或负，以便穷尽可能（详见4.3.二元属性系统节）。还有一类特征标记是单元的标记法，即给出一个无值的标识符，直陈所需特征。如下面几个例子：

[第三人称（3rd person）]	[时态（Tense）]
[主格（Subject Case）]	[宾格（Object Case）]
[所有格（possessive Case）]	[间接格（Oblique Case）]
[施事（agent）]	[受事（patient）]
[经验者（experiencer）]	[述题（theme）]
[目标（goal）]	[受益者（benifective）]
[方位（location）]	[论元（argument）]
[谓词（predicate）]	[嫁接（adjunct）] 等。

一般标识符与特征标识符放在一起可形成新的标识符，如名词性标识符N加上特征标识符 [3rd person，+male，Subject Case，agent，argument，John][①] 就可以在下句话中的主语位置上找到：

[John] likes Mary.

空范畴（Empty Categories）标识符。西方现代语言学理论十分重视空范畴研究。在许多描写式中都出现了空范畴[②]，记作e，t，GAP，pro，PRO等。如，

(a) 张三不喜欢 e

(b) 这书，我不喜欢 t.

(c) 我就是那个你想找GAP，没找着t的好人。

(d) pro 不喜欢我

(e) 张三答应我 PRO 去

所有这些句子可以看成是简化了的结构描写式，可作源描写式，也可作目标描写式。至于说，这些空范畴有什么具体含义及是怎样进入这写描写式中的，那是推导运算上的问题，我们将在下面谈到。

① 这些标示符中，[Case] 表示名词与动词、介词、时态等成分之间的结构关系，与格语法中的语义意义上的“格”不是同一概念。[argument] 是逻辑语义概念，与“谓词”一语相对使用。

② 空范畴的意义和出现原因参见4.7空范畴一节。

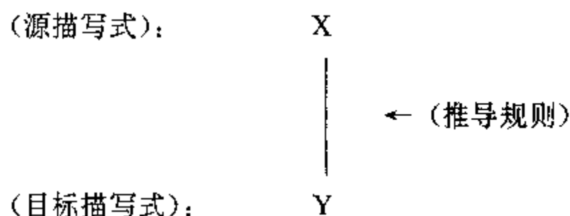
还有一个很有意义的概念也常常用在结构描写式中，这就是标识符之间的距离，有的像(a) 中的"喜欢"与"这种人"那样，离得很近，有的像 (b) 中的"不喜欢"与"这种人"那样，离得很远：

(a) 我 [不喜欢] [这种人]
(b) [这种人]，李四以为张三说过我根本就 [不喜欢]

总之，西方现代语言理论中的结构描写式使用了许多形式化的标识符号，具有浓重的自然科学理论表述的工具色彩。这些符号不是孤立的符号标记，而是语法系统体系中的表述单位，只有在特定的系统中才能理解，才有意义。

一条条孤立的规则只是在语法运算系统中某一环节上的结构体描述，是静态的数据性的表征 (representation)，而不是运算 (computation)、推导 (derivation) 的过程。规则与规则之间必须靠规则运算推导才能建立起必要的联系，才能有机地构成一个具有操作性的运算系统。因此，规则制定的另一个原则是：

一个给定的规则和另一个给定的规则之间的关系叫运算或推导。运算或推导必须作用于一个给定的结构描写式（源描写式）进而导致另一个结构描写式（目标描写式）的出现。推导规则都应符合下面的形式：



因而，规则间的推导关系和推导作用离不开规则表征的需要。比如说，语法系统需要我们从描写式 (a) 推导出描写式 (b)，

(a) A B C
(b) C A B

我们就可以规定和使用下面这个最简洁的推导规则实现：

(c) A B C→C A B

式中的标识符"→"表示把其左边的符号串改写成右边的符号串。这就是人们常见的"改写运算规则"(rewriting rule)，是西方现代语言理论语法系统中极为常见的运算规则。下面一些运算规则也十分常见：

插入规则 (insertion)：其作用是可把描写式 (a) 推导成描写式 (b)：

(a) A　B　C
(b) A　B　D　C

这个规则可以表述成：

A　B　C　→　A　B　D　C

这个规则的作用是在 A B C 这套符号串中加上了一个 C。

删除规则（deletion）：其作用是把描写式（c）推导成描写式（d）

(c) A　B　C　D
(d) A　B　D

删除规则可以表述成：

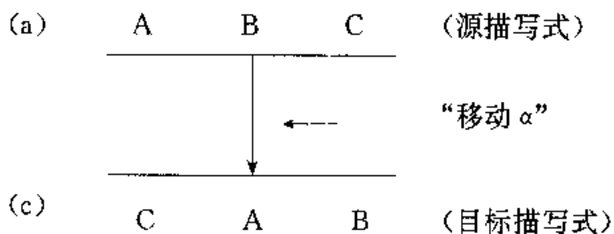
A　B　C　D　→　A　B　D

这个规则的作用是把符号串 A B C D 中的 C 删除掉。

以上规则合并联用，写成一条规则叫“转换规则”（transformation rules），这在西方语法系统中极为常见。

这类具有运算作用的规则常见的还有：

(1) “移动 α”（Move α）。根据这个规则，(a) 可推导成（b）：



(c)　C　A　B　（目标描写式）

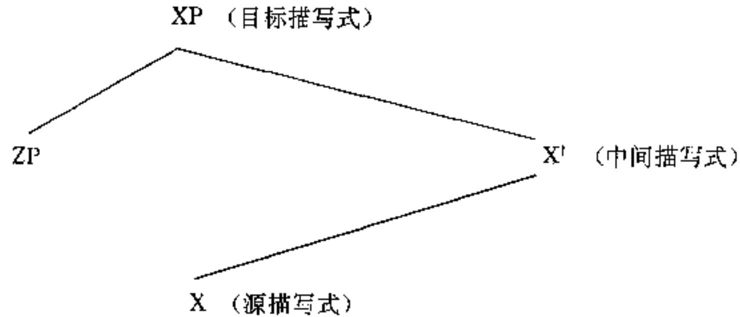
(2) “裹挟移动”（Pied-Piping）。其作用是让一个结构成分随另外一个结构成分移动。例如：

[for whom] [did you solve the problem t]

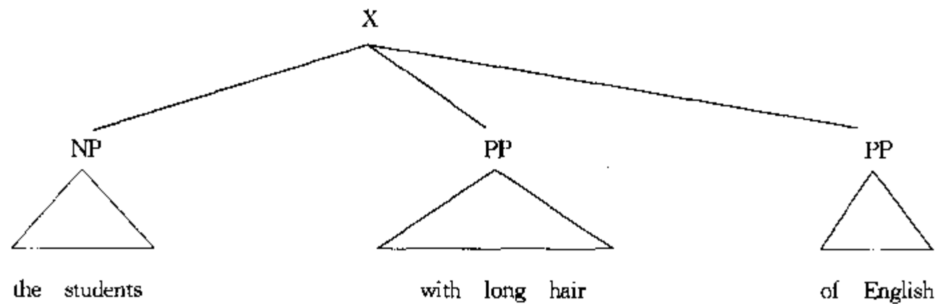
式中的 for 跟随 whom 移动到最前面，构成问句。

(3)“映射”(projection).

这个规则的推导作用是把一个非短语性的结构成分扩展成一个短语性成分。比如，下列推导关系中，X是一个非短语性成分的描写式，可以首先把它扩展成处于XP与X一个中间性的X'，而后再扩展成短语结构XP。



X'结构层次的存在在英语中的证据之一是这样的：英语中的名词性短语“the students with long hair of English”是不成立的，如果我们把它的结构描写看成是下面的样子：

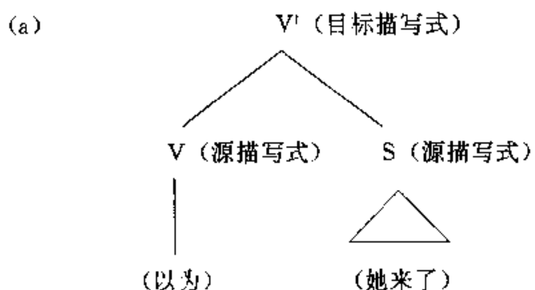


这显然不是我们想要的结果，因为“the students with long hair of English”在英语中是不可以说的。一个有效的把(2)排除的办法是用X'的层次把“students”和它选择的后跟成分“of English”放在它直接管辖的位置上，而让“with long hair”这个不是它选择的成分“嫁接”到N'上。关于这个问题的详细讨论，参见Radford的《转换句法学》(Transformational Syntax)。

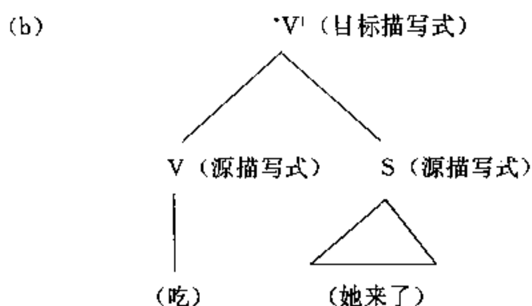
(4)选择或“次范畴结构化”(selection or subcategorization).

这个推导规则的作用是决定一个“中心词项”可以直接后跟一个什么样的短语结构描写式。比如说，“中心词项”V(以为)只选择一个句子描写式跟在它的后面，而“中心词项”V(吃)就不能这样：

(a)

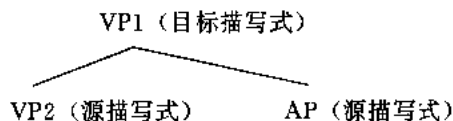


(b)



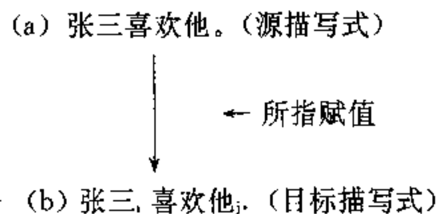
(5) 嫁接 (adjunction)。

只在已有的某些描写式上并入另一个描写式。例如这个推导规则可以把源描写式 AP 接到一个源描写式 VP2 上，得出目标描写式 VP1：



(6) “所指赋值” (assign index)。

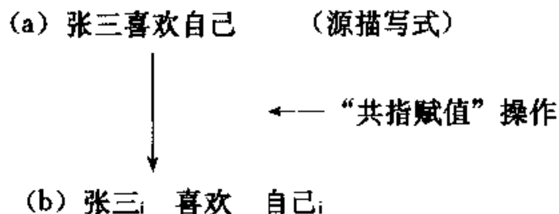
这个操作规则的作用是给某个标识符赋给一个表示所指意义的值。比如说，下面这个运算可以把描写式 (a) 变成描写式 (b)：



“所指赋值”操作把表示指称意义的值 i 赋给了“张三”这个标识符，把另一个值 j 赋给了另一个标识符“他”。

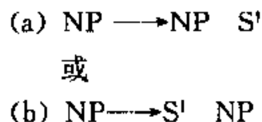
(7)“共指赋值”(coindex).

这个操作的作用是把同一个指称值赋于两个或两个以上的结构成分。比如，下列（a）中“张三”和“自己”就通过“共指赋值”操作成为（b）中的描写，形式化地描写出“张三”和“自己”共指一个人的意义：



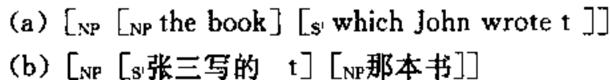
(8)“递归”(recursion)。

这个推导的作用是从一个源描写式导出一个含有这个源描写式的目标描写式。如在处理定语从句现象时常用这一操作：



(a) NP ⟶ NP　S'
或
(b) NP ⟶ S'　NP

执行（a）可得出（a'），执行（b）可得出（b'）：



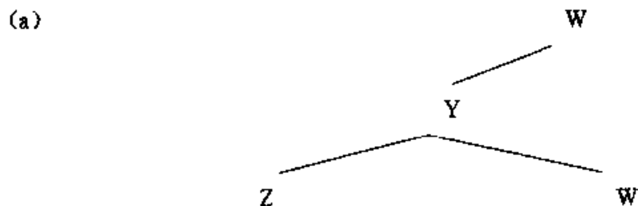
(a) [$_{NP}$ [$_{NP}$ the book] [$_{S'}$ which John wrote t]]
(b) [$_{NP}$ [$_{S'}$张三写的　t] [$_{NP}$那本书]]

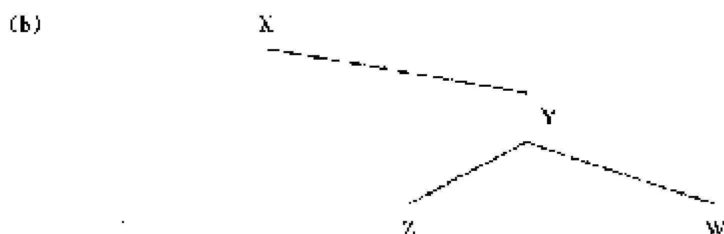
规则的形式化的标识不仅表现在上述的几种情况，还可以从以下几个方面做出规定。

(1)“方向性”(directionality)。

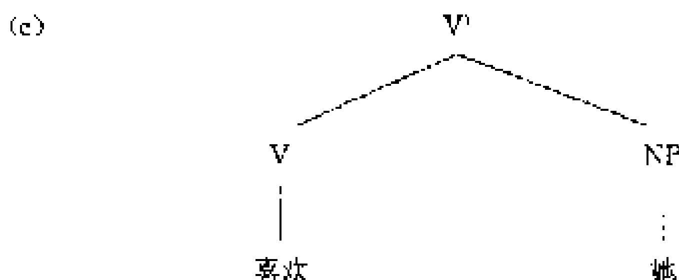
有的结构向左分叉（见 a），有的结构向右分叉（见 b）：

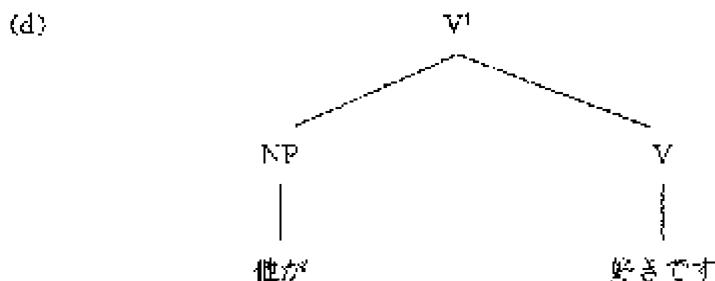
(a)





有的中心词项在候补成分之右（见c），有的中心词项在候补成分之左（见d）：





上面讲到的递归可在左，也可能发生在右。

操作、推导还可分成显性的（overt）和隐性的（covert）两种。比如说，英语中的“疑问词”移动是显性的，而汉语中的“疑问词”移动是隐性的：

(a) [What] [did John buy t]?

(b) [什么] [张三买 t 了呢]?

虽然汉语中不说“什么，张三买了呢？”这样的句子，但是，“张三买什么了呢”这句话中“什么”的语义解释域与英语“what did John buy?”中的“what”相同。为了把汉语中的“什么”的语义解释域像英语中的“what”那样表示出来，给出了这样的结构描写式。这个结构描写式是从源结构式“张三买了什么呢”中通过隐性移动“什么”得到的①

① 西方现代语言学中曾遇到特殊疑问句者或是“求解”的过程，比如说，问“张三昨天买了什么”就是要听话者在“张三昨天所买的东西”中找出各个东西的名字，这样“张三昨天所买的东西”就成了疑问词“什么”的语义域，表示这个语义域的区间大小，语言学中采用了“一阶谓词演算”形式。参见§5比较研究一节。

以上讲的大多数都是有关句法的规则。语音部分规则系统的建立基本方法同句法，由于其中涉及到许多“音位学”上的概念和术语，就不在这里提及。

总之，在西方现代语言理论中各家都有自己的一整套操作运算或推导的规则。有的术语不同，规则内容相同，有的规则条件不同，有的是自家杜撰的规则。但是，每个规则都是系统中的规则，离开了系统，规则就失去了原本的意义。我们认识规则也好，借用规则也好，都必须在其所处的系统意义上认识和使用。离开了原来的系统，即使同一个名字的规则，也失去了原有的意义，只能重新在它所依存的系统中获得新意。东一家，西一家地借来一些规则拼凑在一起，不会构成一个有机的运算推导系统。

语法系统有了足够的结构描写式和推导结构描写式的规则，还不足以保证最终生成全部可能的合乎语法的句子，同时保证不会生成任何不合乎语法的句子。道理很简单：因为所有的操作规则还缺少在什么情况下可以使用，在什么情况下不可以使用的规定，因而一个推导、一个操作便会变得十分任意，就不可避免地生成一些不合乎语法的句子，即关于不合乎语法的句子的结构描写式。比如说，执行一个不含条件的“移动 α”的运算，可以得出（a）这个合乎语法的句子，也可以得出（b）这个不合乎语法的句子：

（a） 我们吃饭的饭馆很漂亮①

（b）*我们吃饭的朋友很漂亮

显然，操作规则的执行需要条件的限制。在所有的西方现代语言理论中，在制定规则的同时，寻找和制定规则的使用条件一直是语言学家的最大努力所在，他们也确实找到了不少成功的规则条件。这些规则条件的一个共同特点是：条件中所使用的限制性术语、概念都是在给定的结构描写式中已有的形态构件之上定义的。因此，规则的使用条件是系统之内的条件，也只在系统之内才有意义。下面选出几个比较有名的规则条件说明。一个在整个西方现代语言理论界影响很大的规则条件叫“结构岛条件”（Island Conditions）。这个条件用来限定“移动 α”规则的使用。它的大致规定是：在一个结构描写式中某些成分会结合成一个像“孤岛”一样的成份，任何一个出在这个“孤岛”上的成分都不能从这个“孤岛”中脱身，移动到前面或后面去。否则就会得到一个不合乎语法的句子。这种“结构岛”至少有三种：（1）“NP 岛”（NP-island）；（2）“wh 问句岛”（wh-island）；（3）“嫁接成分岛”（adjunct island）。体现这些规则条件的例子分别见于（a）、（b）和（c）：

（a） * ［［NP t 写的那本书的］［那个人］］来了

（b） * ［what］do you think ［where John bought t］

（c） * ［这本书］［［如果你要 t］我就要那本书］

---

① 证明（a）中有“移动”发生的办法可以是这样的：（a）“的”字肯定在［我们吃饭］这个句子之外，因为在汉语的任何一句话结尾处都可以加上一个“的”字。反过来说，任何一个可以加上“的”的结构（除名词和形容词外）都是句子。（b）［饭馆］又肯定是从这个“的”字前面的句子中出来的，不然的话就应允许下面这样的结构存在：*［［我们在家里吃饭］的饭馆］。把（a）和（b）逻辑地联系起来就可以断定这句话中有某种“移动”发生。

(a) 句中，[那个人] 从 [t 写的那本书] 这个 NP 岛中移了出来，(b) 句中，[what] 从 wh 问句 [where John bought t] 中移出，(c) 句中，[这本书] 从作为条件从句的“嫁接成分”[如果你要 t] 中移出，都违反了不能从结构岛中移出的规定，结果都是不合乎语法的句子。在“结构岛条件”的限制下，含有移动规则及使用这个规则的“结构岛条件”的语法系统就能保证不会生成这样的句子，正好反映了人从来不说这类不合乎语法句子的事实。

在西方现代语言理论中，人们长期讨论的有名的规则条件还有以下几种①：

(1) “毗邻条件” (Subjacency)
(2) “空范畴原理 “ (The Empty Category Principle)
(3) “语障论” (Barriers)
(4) “经简原理” (The Principle of Economy) 等。

总之，西方现代语言理论的语法系统具有以下几个特点：

(1) 语法系统是一个嵌套在认知系统之中的子系统；
(2) 该系统同信念语用系统及语音系统接口；
(3) 该系统的最终总输出是关于所有可能语言中所有合乎语法句子的意义和声音的结构描写式；
(4) 该系统是个由足以用来保证只输出关于所有可能语言中所有合乎语法句子的结构描写式的规则及规则条件组成的运算推导系统，具有鲜明的公理性和操作性。
(5) 该系统中各规则及规则条件完全依据形态结构制定；
(6) 该系统中的规则及规则条件是极其有限的。

要建立这样一个具有这些特点的操作性系统，用 Chomsky 的话来说“需要几代人的努力”，世界各地许多语言学家通过对各种不同语言的研究，正在为建立起这样一个极为复杂的操作系统做不懈的努力。在这种努力中，他们都首先面临这下面几个不得不回答的方法论问题：

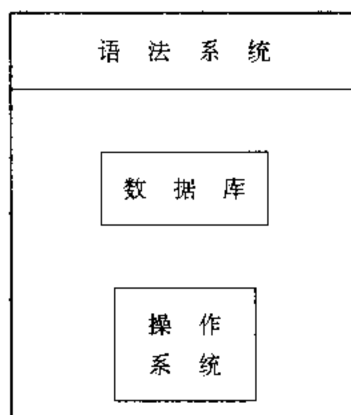
(1) 如何在众多的人类语言的无限多的语言事实中找出极其有限的规则及其条件？
(2) 如何在众多的不同的语言中找到统一普遍的规律和原理？
(3) 如何把孤立的分散的规则及其条件组织成几个简单的系统模块？
(4) 系统中究竟应有几个内模块？每个模块中又有哪些规则？

3.3. 语法系统内模块的建立

西方现代语言理论中，各家语法系统的内模块组织结构有很大的差异。但是，无一例外

① 这些规则条件涉及到许多技术术语、繁杂的推理过程和多种语料。

的都把语法系统的内部结构分成数据库和操作系统两个部份：



这和计算机软件程序的设计思想一样。在所有的西方现代语言理论中，数据库部分都称作“词库”(lexicon)，其中存放着一些称作“词项”(lexical items) 数据型标识符，有些像一般常见词典中的单词，不过这两者之间有很大的差异。词库用比喻的意义讲指的是某种形式的“人脑词典”。其中，每个词项都是一个由一些必要的表示语音、词法、句法、语义等特征的标识符构成的“集合”，而不是常见词典意义上的单词。比如相应于常见词典“喝”一词的词项应该是下面的样子：

相应于常见词典“喝”一词的词项形态：

[[语音特征： [+velar，aspirate/ −back，−high]]
[句法特征： [−N，+V]]
[次范畴化特征： ______NP]
[语义角色特征： 01 02]
[格位特征： [SUBJ，OBJ]]

……]

在西方不同语法体系中，一个词项究竟由哪些特征来描述，各种描述使用什么术语和标识符有着很大的差别，这是因为各家在对语法系统中的词法与句法之间的接口看法不同造成的。生成语法的前期依据的是“句法推导词法学”(derivational morphology)，把“时体”“人称”等特征单独看成一个词项。而后期，由于依据“曲折变体词法学” (inflectional morphology) 把“时态”“人称”等看成动词的两个不同的特征分别记在动词上。概括性短语结构语法，格语法及生成语法的“扩充式标准理论”使用的语义特征较多，而生成语法的“管约理论”模型所使用的语义特征则十分局限。与常见词典相比，这种词库中的词项多采用了 Frege 的思想，每个词项只有一个意义。这是因为使用了特征集合的方法刻画词项所决定的。另外，许多在常见词典中不是单词的成分也常常被列为词项，上面提到的“时态”“人

称”就是两个例子。像标记从句结构的隐性句法成份（生成语法）[①] 或 XCOMP（词项功能语法），表示某些空范畴的句法、语义成份（如 pro、PRO、GAP 等），表示指称含义的“下标”（sub-index）等。

在操作部分的设计上，各家理论之间以及同一理论的不同发展时期表现出一些差别，有的甚至很大，这不仅表现在技术用语上，也表现在理论内容上。生成语法的“句法结构”理论模型，“扩充式标准理论”模型和“支配管约理论”模型把语法系统作“句法”（syntax）或“广义句法”（broad syntax）。叫“广义句法”的原因是因为这个部分既包括句法又包括“音位操作”（phonology，实为 phonological representations / operations）。关于音位操作部分，几乎所有的语法模型都把它看称是音位学家们可相对独立完成的子系统，在模式中只按一般条件定义处理。词项功能语法中“成分—结构”（constituent-structure，近似于生成语法中的 s-structure）和概括性短语结构语法中的“表层结构”（surface-structure）既担负着与音位部分接口的任务又担负着同信念语用系统接口提供语义解释的任务；而在生成语法中即有一单独与语音系统接口的音位部分，同时还有一个与信念语用系统接口的语义解释系统，句法语义描写与音位描写是各自独立的。不过，在生成语法与概括性短语结构语法和词项功能语法之间还有一个重大的差别：生成语法在与音位部分并列的表层结构（surface-structure，s-structure）之后还有一个逻辑式（Logical Form），表层结构不是最终的语义描写式。在概括式短语结构语法中，最终的语义描写式是表层结构；在词项功能语法中，语义的最终描写式是“功能结构”（functional-structure，f-structure）。造成西方现代语言理论各派之间的这些差别的主要原因有二：一是在立论过程中关于“描写上的充分性“和“解释上的充分性”两者的权重不同。生成语法（中期和近期）和概括性短语结构语法在立论过程中首先追求“解释上的充分性”，即理论解释上的深度，而后扩展描写上的广度，因而在形式化的程度上要求很高。而词项功能语法比较偏重首先追求描写上的充分性，追求更逼近心理上的真实性。

小结：

讨论到目前为止，我们可以这样小结一下西方现代语言学家在建造语法模型系统过程中所表现出的方法论总体特征：

---

① 英语中显性的从句标示符 C 下可出现 “did”、“can”、“that”、“for” 等。汉语中没有相应的显性成分，这样可以看成有相对应的隐性成分。也有人认为汉语有某些标示从句的显性成分，如 “由”、“给”、“吗”、“呢”、“呐” 就可能是其中的五个：

[这件事] 由 [S 我处理]
[这件事] 给 [S 我忘了]
[你忘记了这件事] 吗？
[你忘记了什么] 呢？
[这件事] 吗 [我忘了]

可以看出，这五个成分都和英文中的显性 C 成分一样处在句子之外。

(1)他们首先从人人都有区别合乎语法的句子和不合乎语法的句子的能力这一事实出发，确定了语法系统在整个认知系统中的子系统独立性，从而确定了语法系统的总体输出。

(2) 由于他们所要建立的语法模型是一个具有公理意义的操作性形式化系统，语法系统的总输出被明确地规定为关于所有可能的合乎语法的句子的音位特征的描写式和语义描写式。

(3) 由于他们所要建立的语法体系首先不是关于具体语言的具体语法，而是赖以能够获得任何一种人类可能语言的普遍语法，即一种可以推导出所有人类可能语言具体语法系统的普遍语法系统。

(4) 这个普遍语法系统的内模块仍然是一个操作性的形式化系统。这个系统和任何一种操作系统一样具有从输入到输出的操作运算性，可把一个表征描写式推导成另一个表征描写式。

(5) 全部系统由表征规则和运算推导规则及规则使用条件构成。

要具体实现这种严格意义上的操作性系统，技术上的关键可以归结到一个问题。这就是如何用十分有限的规则及规则条件推导出无限的关于合乎语法的句子的结构描写式来。用比喻来讲，这就好像语言学家有了一个总任务，剩下的就是一个实现这一总任务的具体分工问题了 (linguistic division of labor)。选择最合适、最经济的分工便成了推动语言理论发展的推动力和语言学家们的追求。

3.4. 语法系统内模块建立的普遍语法原则

语法系统内模块建立遵循的基本原则叫普遍语法的原则。

前面已经讲到，西方现代语言理论所共同追求的目标是寻找普遍语法的东西，是建立一个普遍语法系统，从这个系统中可以推导出所有可能语言中的所有可能的句子来。但是如何能在众多的情况各异的具体语言中找到普遍遵循的规律，如何在具体语言中无限多的语言事实中归纳出数量有限的规则来？

西方现代理论语言学和所有语言理论一样都承认语言的社会属性，语言间和方言间的差异以及同一语言（方言）使用者之间的个体差异。但是，这些差异并不是语言的本质方面，并不反映语言与非语言之间的本质差别。这就和张三和李四之间的差异不反映人和其他动物之间差异一样。研究一种语言与另一种语言的差别，研究这个社团语言与那个社团语言的差别，研究同一语言不同发展历史时期的变化无疑都是具有科学意义的工作。由于，西方现代理论语言学的研究兴趣不在语言的这些个体差异方面，也不在语言的社会属性方面，而在于从人与动物的区别意义上考察研究语言，找出语言的自身系统性，所以在总的研究方法上强调使用“理想化条件研究方法”(idealization)。这种研究方法具体体现在以下几个方面：

第一种可称作“规则形式限定法”。其主要内容是：语法系统中的规则全都取统一的操作运算方式。比如说，统一的“改写规则形式”：

$$X \rightarrow Y \quad Z$$

这是人们在60年代前后在转换生成语法中常见的改写规则。所有具体语言的具体语法规则都必须表达成这种统一的形式，每种具体语法便都是由用改写规则写成的规则所组成的规则集合。比如说，语言 A 的语法 GRAMMAR-A 就是由规则 R1，R2，R3，……RK 组成的规则集合；语言 B 的语法 GRAMMAR-B 就是由规则 RR1，RR2，RR3，……RRK 组成的规则集合；语言 C 的语法 GRAMMAR-C 就是由则 RRR1，RRR2，RRR3，……RRRK 组成的规则集合：

GRAMMAR-A {R1，R2，R3，...... RK}
GRAMMAR-B {RR1，RR2，RR3，...... RRK}
GRAMMAR-C {RRR1，RRR2，RRR3，...... RRRK}

在所有这些规则集合中，每个规则都必须按照统一的改写规则形式书写，凡是按改写规则形式书写的规则都会生成合乎语法的句子，凡是不按改写规则生成的句子都看作不合乎语法的句子。这样，可能语言中的全部可能句都就范于改写规则的形式上。举例说，早年在 Chomsky 的“语法结构”中就出现了下面一组改写规则：

i. S → NP (AUX) VP
ii. NP→ (D) (A *) N
iii. VP→ (ADV) V (NP)

（以下从略）

执行这套规则，我们可以得出以下这类的汉语句子：

(1) [S [NP 张三] [VP 来了]]
(2) [S [NP 张三] [AUX 不能] [VP 来]]
(3) [S [NP [D 那] [N 孩子]] [VP 来了]]
(4) [S [NP 张三] [VP [V 刷] [NP 张三]]

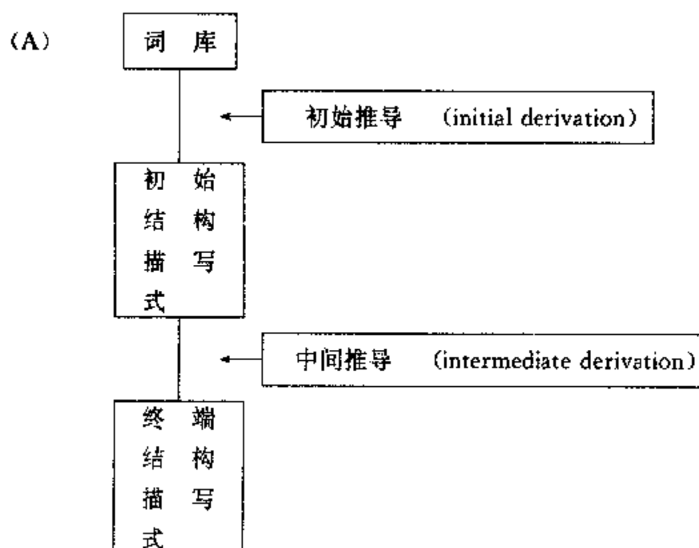
也可以生成像下面这些的英语句子：

(5) [S [NP John] [VP came]]
(6) [S [NP John] [AUX will] [VP come]]
(7) [S [NP [D the] [N boy]] [VP came]]
(8) [S [NP John] [VP [V brushes] [NP John]]

但不会生成下面这样的句子：

(9) * [S [VP 来了] [NP 张三]]
(10) * [S [VP came] [NP John]]
(11) * [S [VP 来] [AUX 不能] [NP 张三]]
(12) * [S [VP come] [AUX will] [NP John]]

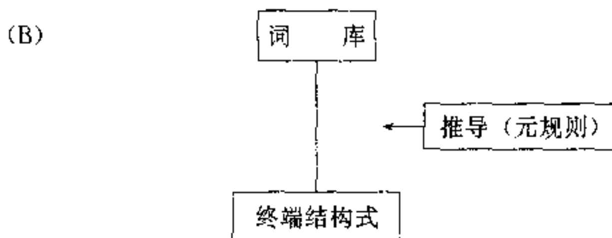
按照这个思路组织起来的语法系统确实有普遍语法的限制作用，可以想像出来：这种限定确实可以保证得到所有语言中的所有可能的句子。但是，这种限制显然过松，它所生成的句子还可以进一步得到限制，避免出现所谓“过剩生成”(over-generation) 的问题。这不仅在原系统设计预料之中，也是建立这种庞杂的语言系统必要的理论过程。正是由于用规则形式来限定人类语言的可能性有“过剩生成”的弊病，语言学家们进而开始把普遍语法限制从规则形式上转到规则使用条件上。这就引出了第二种组建普遍语法规则的方法：“规则条件限定法”。这种方法的基本内容是：同一规则的使用条件在各种语言中都是一样的。比如说，“A-OVER-A 条件”(A-OVER-A Condition)，“特定主语条件”(Specified Subject Condition)，“时态句条件”(Tensed Clause Condition) 等，就是有名的规则条件限制，在西方现代理论语言学各家著述中常见。虽说这种规则使用条件有很强的普遍语限制作用，但是仍旧十分零散，缺乏理想的普遍语法性，因为规则总是具体语言的具体规则，本身并没有超越具体语言特异处的概括性。语言学家们便开始从两个新的方向寻找更具有普遍语法意义的东西。在语法系统的模块中找出一个超越语言特异处的部分，作为操作系统的最初始的一个结构描写式。这就是人们熟悉的“深层结构”(deep-structure)。深层结构是最初始的操作规则从数据型的词库中推导出来的初始结构描写，在理想的操作性系统中，作为初始结构描写式的深层结构应该在数量上极为有限，而后便可通过另外一些有限的操作推导，产生无限多的结构描写式，以满足系统总输出的需要。这个思想可用 (A) 中的框图表示。



用于初始推导和中间推导的操作性规则都是有条件的，这些条件都是使用在词库中已定义了的特征标识符或/和结构式中已有的结构标识符来表述。但是，由于人们在某些标识符应该出

现在什么地方，应该出现在词库中的词项上，还是应该出现在初始结构描写式中，还是应该出现在终端结构描写式中，看法不一样，便导致出现了不同的内模块结构。比如，概括性短语结构语法认为表征的初始结构描写式应用操作性的推导规则来代替，这样一来，初始结构描写式连同初始推导规则就全都变成了终端结构描写式和中间性推导，也就取消了初始结构和初始推导与终端结构和中间推导的界限，系统内部结构就变成了下面的样子：

(B)



同（A）相比，这个系统组织结构模块看上去要简单。因为去掉了中间的推导环节，但是，原来在（A）中由初始推导和初始结构承担完成的任务现在得由统一的推导规则来完成。这样，原有的限制在初始部分的规则条件就失去了其限定的作用，因为规则使用的环境已经改变，当把这些初始规则同中间规则放在一起在一个环境中使用时，就不得不增加新的规则条件，才能确保规则条件的限定作用。从（A）和（B）之间的比较，可以看出西方现代理论语言学家们调试语法系统内模块的思想方法。这个思想路线可以这样描述：假定一个操作性的语法系统总任务为 T，为实现这个总任务的规则为 {R1，R2，R3，……RK}，规则条件为 {C1，C2，C3，……CK}，定义规则条件所用的标识符为 {F1，F2，F3，……FK }，词项为 [I1，I2，I3，……IK}，那么就可以以标识符 {F1，F2，F3，……FK} 为出发点，有三个去处可以把它们分派出去：一是词项身上，一是表征结构描写式上，还有规则使用条件本身。总的看来，

（1）普遍性的、原理性的东西用规则（A）初始结构和（B）规则条件实现。具体语言中的特异部分（idiosyncrasies）用规则来实现。规则仍是因语言而异。

（2）普遍性的、原理性的东西用加在词项上的条件来实现。

（3）普遍性的东西统一在描写式中表达。

这是 60 年代到 70 年代，美国理论语言学界各家设计语法系统内模块的基本考虑。与此同时，人们开始把多个规则归并抽象成一个简单的规则，用“移动规则”代替繁多的转换规则就是最有名的例子。而后，人们开始用普遍语法的原理代替只适用于具体语言的规则。把具体语言间的差异用同普遍语法原理相连的参数变化来实现。这就是普遍认为最成功的“原理—参数法”（Principle-Parameter Approach）。

70 年代，人们发现因语言不同而相异的规则虽说有很充分的描写力，但缺乏足够的理论解释力，因而 Chomsky 首次提出了有名的“原理—参数法”。开始用原理代替规则以获得必要的理论上的解释力，用参数对原理的作用保证事实描写上的充分性。所谓原理是在语言特异的规则基础之上概括抽象出来的具有普遍意义的原则性法规。它的主要特征是不拘泥于某一语言的一种语言事实或一类语言事实，而是对多种语言间，多种语言现象间的形式化概括。

比如说，把过去的“结构岛条件”、“A-OVER-A 条件”、“特指主语条件”（Specified Subject Condition）和“时态句条件”（Tensed Sentence Condition）等就归结成一条“毗邻原理”（the Subjacency Principle）。这一原理具有普遍语法意义[①]，在任何一种语言中都应实用。但是，原理中所使用的“临界结”（bounding nodes）的值，即什么成分是一个“临界结”要因语言的不同而不同，这样就规定出了极为有限的几个参数同这个原理相联，当参数被确定下来之后，这个原理就会起作用。可以说，70 年代后，西方理论语言学学家们把极大的兴趣和精力都投入在寻找普遍语法原理及其相关参数上了。这个“原理—参数法”一直延续到今天，是西方当代理论语言学界的一个主导思想方法和理论方法。

① 而后，又把这些原理连同“外移域条件”（Condition on Extraction Domain）及“空范畴原理”等归纳成了“语障论”（Barriers）。现在又试图把这些归纳成更抽象的“经简原理”（The Principle of Economy）。

4. 西方现代理论语言学方法论的具体特征

西方现代理论语言学作为公理性的操作系统在理论上除了要满足两个充分性：事实描写上的充分性和理论解释上的充分性外，还必须满足一切科学理论都追求的预见性和简洁性，实现这些要求的具体方法有：

(1) 对比研究法
(2) 理论逼近法
(3) 比较研究法

4.1. 对比研究法

4.1.1. 在可能与不可能之间寻找限制

语言研究者所面对的语言事实是无限的，而语言研究者的任务是要找出造成无限事实的有限的成因来。西方当代理论语言学家讲得更加明确，不但要找到一种语言中无限语言事实的成因，而且要找到所有可能语言中无限语言事实的成因。实现这个理论目标，简单的事实归纳法显然是无济于事的。为了实现在无限的事实中寻找极其有限的限制，西方当代理论语言学学家们首先认为人类的语言不是任意约定之物，而是有限制的。那么，如何在无限事实面前寻找有限的限制呢？他们认为限制在可能与不可能之间，在是什么与不是什么的差别之间，从而开始了在可能与不可能之间，在是与非之间寻找限制的长期努力。

忽略方言间的差异和方言与语言间的异同不计，世界上史前曾有10 000到15 000种人类语言，现有6 000种之多。如果说，一种语言是一套人们用于交际的符号系统，那么现在世界上就会有6 000种之多的用于人类交际的符号系统（当然，语言不是唯一的用于交际的符号系统）。但是，不管人类语言数量有多么多，不可能任何一种符号系统都可以成为人类语言。这就和人类的音乐系统一样，不是任意一种声音系统都能成为人的音乐系统。比如说，人类的音乐系统只能建造在1、2、3、4、5、6、7这七个音符之上。任何一个含有这七个音符之外的声音系统都不会成为人类音乐系统。因此，什么可以是人类音乐系统，什么不可能是人类音乐系统便有了限制。这种限制不可能是人为的，也不会是约定俗成的。无论人们怎样约定，下面的这个含有“8”这个人造音符的东西决不会成为一个可被人来吟唱的乐句：

*8　1/2　2/3　8/1 — /

因此，这就和一个不合乎语法的句子一样是一个不合乎人类的“音乐语法”的乐句。人类的音乐系统不管有多么复杂，乐曲种类有多么繁多，用乐句构成的乐曲篇章多么变幻无穷，有一点是确定无疑的，那就是，这个音乐系统不含有 8 这个标识符。音乐系统是这样，语言系统也应如此。比如说，不管我们怎样约定，不管历史上怎样约定过，将来又会怎样去俗成，讲汉语的人也不会讲出下面这些句子：

* 张三来了李四
* 我以为她
* 你喜不喜欢她为什么不来

讲英语的人也决不会讲出下面这些与上面那些汉语相对应的句子：

* John came Mary
* I think him
* Do you like why she didn't come

讲法语的人决不会讲出下面这些与上面汉语或英文相对应的句子：

* Zhangsan est venu Lisi.
* Je'll crois.
* Est-ce que tu l'aimes pourquoi elle n'est pas venue?

这些不合乎语法的句子就和那句不合乎音乐语法的乐句一样，说明语言是有限制的，超出了固有的限制就没有人类语言可言。那么这些在许多（可能是所有人类语言）语言中都是不合乎语法的句子向我们揭示了什么限制呢？只凭对上面这些简单的语言事实的观察，我们可以得出这样的结论：在所有语言中什么成分可以跟在什么样的动词后面是有一定规矩的。比如说，任何成分都不能跟在像“来”及在其他语言中与其意义相对应的词语后面。名词不可以跟随在“以为”和它在其他语言中与其意义相对的词语后面。另外，从以上三种语言中的第三句话中可以看出：一个“特殊疑问句”不可以出现在一个“一般疑问句”（Yes-No Question）中。把这些观察用形式化的语言联同从对其他一些语言事实的观察中得出的结论有机地放在同一个系统中，这个系统就足可以限定了什么可以是人类语言的可能性。

4.1.2. 对比研究法

可是，这里却掩盖了一个人们常常会忽视的问题：凭什么当我们看到“张三来了李四”这句话就得出结论说“来了”这个动词后面不能跟名词呢？对于这个问题的回答似乎十分简单：因为我们从来不说“张三来了李四”。实际上这里暗含着一个重要的判断过程，那就是一个“对比”的过程：因为我们从来不说“张三来了李四”而只说“张三来了”。这就是说，在下面的对比中得出的结论来的：

(1)　张三来了

(2) * 张三来了李四

一个会讲话的人是这样凭直觉区分什么能说，什么不能说，反过来，从什么能说，什么不能说，什么是句子，什么不是句子的分析中我们可以认识这个人的直觉内容。而这种关于什么是句子，什么不是句子的语言直觉正是语言学家所要寻找的什么是可能语言的限制。这样，旨在寻找限制的西方现代理论语言学家便采用了各种同语言使用者在判断句子与非句子时所用的一样的方法，叫作“对比法研究”(contrastive study)。

“对比研究法”是西方现代语言学的一个最基本的具体研究方法，在各个理论模型中得到广泛的应用。下面讨论一下这种方法应用上的一些具体技术问题。

对比的形成不是任意的。上面的 (1)、(2) 两句可以形成对比，但是 (3)、(4) 之间不能形成对比，至少不能形成有理论意义的对比：

(3)　昨天下雨了

(4) * 张三来了李四

而 (5)、(6) 之间也不会形成有意义的对比：

(5)　张三昨天来了

(6) * 张三来了李四

(3) 和 (4) 之间不能形成对比的原因是，两者之间没有相比的基础，两句之间没有共同之处。而 (1)、(2) 之间除了 (2) 句句尾出现了“李四”而 (1) 句句尾没有“李四”，也没有任何其他成分之外，其他的地方全部相同：

(1)　张三 来了

(2) * 张三 来了 李四

这种“只有一点不同而其余全部相同”的两个句子或其他序列符号称作“最小差异对”(minimum pair)。(5)、(6) 之间虽说有共同之处，但它们的差异却不只一处：

(5)　张三 昨天 来了

(6) * 张三　　来了 李四

因此，(5)、(6) 不是一个“最小差异对”。按照这个道理，(3)、(4) 也不能被看作是“最小差异对”，因为它们之间处处相异。(7)、(8) 也不会构成一个“最小差异对”：

(7) * 张三 来了 一本书

(8) * 张三 来了 李四

这是因为，两句话都是不能说的汉语，两者之间没有任何可比之处。但（9）、（10）是一个“最小差异对”：

(9)　张三 看见了 李四
(10) * 张三 来了 李四

与（1）、（2）之间相比，（9）、（10）之间也满足了“只有一点不同，其余全部相同”的条件，不过两者的不同处不一样：（1）、（2）不同处在于句尾是否有“李四”一词，而（9）、（10）之间的不同之处在于它们在同一个位置上使用了不同的动词，（9）用的是“看见了”，（10）用的是“来了”。至此，我们找到了两个“最小差异对”，分别陈列如下：

(11) a.　张三 来了
　　b. * 张三 来了 李四
(12) a.　张三 看见了 李四
　　b. * 张三 来了 李四

自然识别或有意构成“最小差异对”对于本族语使用者和语言研究者来说都不是困难的事，因为这种“最小差异”本来就存在于语言使用者的语言直觉知识之中，他们不需要学习也不需要专门训练就能够辨认出差异来，语言研究者如果本人就是所及语言的使用者，他自身的语感和任何一个同一本族语使用者的语感一样，有着天然的识别差异的本领。语言研究者凭自己的语感有意构造出一个最小差异对和作为语言使用者鉴别一个最小差异对，不管从过程上看还是从结论上看都是一样的。但是，对于语言研究来说，意义不在识别最小差异对，而在于如何解释，在于解释像（11）、（12）这种最小差异对说明了什么问题。以（11）、（12）为例。假定我们像（13）这样对（11）中的最小差异对作以下理论说明：

(13)“来了”这类动词后面不能跟有名词。

这似乎足以解释为什么（11a）合乎汉语语法，（11b）不合乎汉语语法。但是，这一结论却不能排除我们可以说出（14）这类不能说的汉语的可能：

(14) a. * 张三来了聪明
　　b. * 张三来了李四不认识我

（14）中的事实逼迫我们不得不修改（13）中的观察和解释，因为（14a）和（14b）中，动词“来了”后面跟的不是名词而是形容词“聪明”和句子“李四不认识我”，这并没有违反（13）中的规定。所以，（13）似乎应该写成（15）的样子：

(15)“来了”这类动词后面不能跟有任何成分。

但是这个结论又不能解释为什么可以说(16)这样的句子：

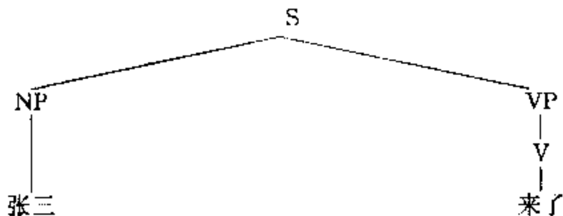
(16)张三来了两次

因为“来了”后面跟着“两次”。按照(15)所规定的道理，(16)应该是不合乎汉语语法的。究竟如何在这些最小差异对中找出更理想的解释呢？在西方理论语言学文献中有一个常见的方法是用“次范畴化”或“选择限制”来解决。这种处理方法的大致意思是：所有动词的后面能跟什么，不能跟什么都有一定的限制，而每个动词都必须遵循这些限制。一种形式化的表述这种限制的方法是：

(17) a.“来”　/______0
　　b.“看见”/______NP
　　c.“以为”/______S　　　等。

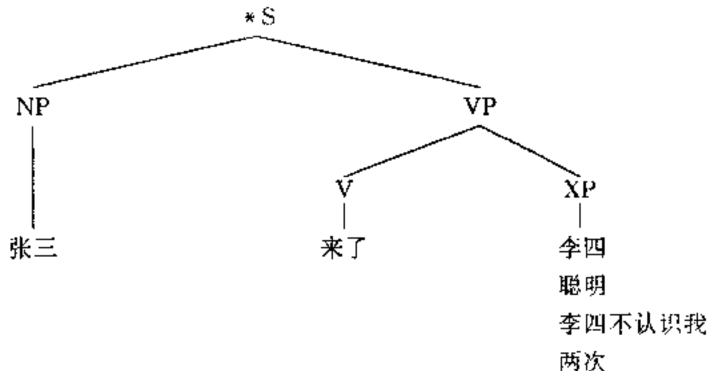
(17a)表示“来”只能出现在什么成分都没有的环境下；(17b)表示“看见”应出现在一个名词短语之前的语境中，(17c)表示“以为”应出现在句子之前的语境中。这些信息都是在词库中规定好了的，当动词“来了”进入句法操作时，就呈现为下面的结构描写式：

(18)



这个规定是强制性(obligatory)的，就是说只能这样别无选择。既然如此，(19)便自然成为不可能的了：

(19)

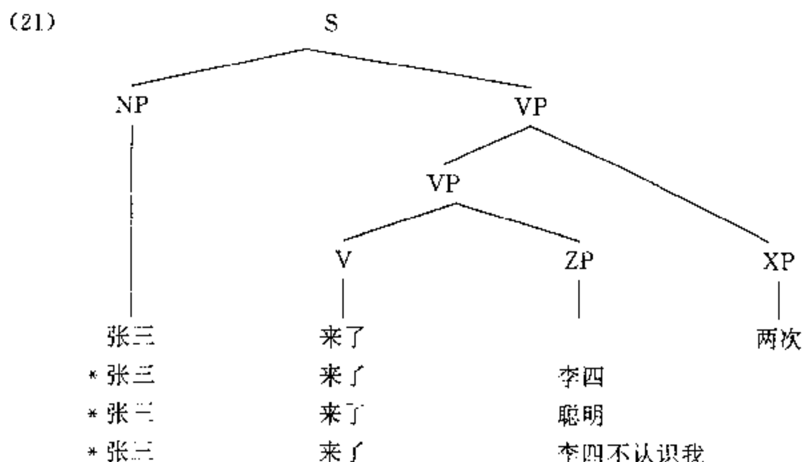


但是，这却连（16）这样可以说的话也给排除掉了。问题出在哪里呢？问题出在“跟在后面”一语上。这就迫使我们开始注意另外一个最小差异对，这个最小差异对表现在（20）里：

(20) a. 张三 来了 两次
b. * 张三 来了 李四
* 张三 来了 聪明
* 张三 来了 李四不认识我

可以想到的一个理论解释是：“两次”不跟在“来了”的后面，“李四”、“聪明”和“李四不认识我”跟在“来了”的后面。这里的最小差异对很可能在下面的结构描写式中得到解释：

(21)



说一个成份“跟在另一个成份的后面”或不“跟在另一个成份的后面”，除非在特殊的定义下，一般指的是一种线性的排列关系，因此“来了两次”和“来了李四”在这种意义上是没有差别的。不过，在（21）中，“两次”是“嫁接”在 VP 之上的，而“李四”、“聪明”和“李四不认识我”都是处在“来了”的“补语”位置上。有了这种关于“两次”和“李四”等成分的句法结构上的差别，我们就可以精确地解释（20）中的对比：

(22)“来了”这类动词只能选择 0（即什么东西也没有）为补语（补语是一个严格结构概念，只能指相关动词所处短语结构的像 ZP 所处的位置）。

由于“来了”不能选用任何补语，而在（20）中却分别选用“李四”、“聪明”和“李四不认识我”，结果给出了不合法的句子，即人们从来不说的非句子。另一方面，由于“两次”不是“来了”选用的结果，而是在它选用 0 后添加上的，因而，“张三来了两次”是一句合乎汉语语法的句子。到此为止，我们算讨论完了（11）这个最小差异对，但还没有触及到（12）这个最小差异对的问题。为方便，我们把（12）再重复写在下面：

(12) a. 张三 看见了 李四

b. * 张三 来了 李四

说“来了”这类动词不可以选用补语，仍不能确定这是不是这类动词的本质属性。如果其他类或所有动词都像“来”这类动词一样不可选用补语，那么“不能选用补语”也就不能成为这类动词的本质。(12) 这个最小差异对恰好说明了这个问题。当动词“看见了”在 (12a) 中那样选用了一个补语“李四”时，其结果是一个合乎汉语语法的句子，相比之下，(12b) 由于“来了”选用了同样一个补语“李四”，结果生成了一个不合乎汉语语法的句子，对比的确说明，“来了”代表着一类动词，这类动词不能带补语。但是，这仍旧不是讨论的完结，因为我们还可以说 (23) 这类句子：

(23) 张三 看见了

结果，刚才得出的关于不可带补语是“来了”这类动词区别于其他动词的本质特征的结论，在 (23) 面前有模糊了起来，因为它们之间的对比消失了：

(24) a. 张三来了
b. 张三看见了

为了解决这个问题，首先，我们有必要想一想，(24b) 中的“看见了”真的就没有选用补语。对这个问题作出判断，我们仍然可以使用“对比法”。在西方理论语言学文献中就有人依据下面的最小差异对鉴别“看见了”在 (20b) 的情况下是否有补语的：

(25) a. 这个人，张三看见了
b. * 这个人，张三看见了李四
c. 张三 看见了李四

(25) 中的对比说明了这样问题：当“话题”(topic)“这个人”出现的时候，“看见了”本可带有的补语不能出现，要么保留话题“这个人”(如 25a)，要么保留补语“李四”(如 25c)，两者不可兼有 (如 25b)。而既没有话题又没有补语的 (24b)，则要么是省略了话题，要么是省略了补语，但省略不等于本来没有。如 (26) 所示，“张三来了”本来就没有补语，也不会有话题：

(26) a. * 这个人，张三来了
b. * 张三来了李四

上面我们通过一些实例，说明了对比法应用的一些基本情况。概括地讲，对比法在最小差异对中进行的。下面讨论一下对比法的理论意义、应用范围和使用技巧上的一些问题。

对比法首先是一种在无限的事实中寻找本质特征的可靠方法。由于语言事实的无限性，对比法便表现出其他方法不可代替的科学理论意义。比如说，我们想知道一种语言中的句子到

底应该是什么样子，到底有哪些本质特征，句子的生成到底遵循哪些条件等。满足这种理论需要的一种传统的方法是把事实搜集起来，从所搜集到的事实中归纳、整理出一些概括性的规律来，这种方法虽不失其科学理论意义，但免不了要碰到这样一个经验性的问题，这就是，事实是无限多的，我们是没有办法把事实全部搜集起来供我们研究，而只能是无限多的事实中的一部分。从无限中的一部分得出来的结论免不了会同从无限中的另一部分中得出来得结论相矛盾。旨在为建立一个可保证生成无限多事实的语言系统的西方现代理论语言学，似乎不能依赖穷举事实的“穷尽法”，也不能依赖“归纳法”。这里，我们不免联想起语言研究之外的例子。比如说，我们想知道人的种属属性是什么，就是说什么使人成为人，就和我们想知道作为句子集合的语言中的句子的属性是什么，就是说，什么使句子成为句子一样。实现这个想法的方法有两种。第一种方法是这样的：我们把张三、李四、王五、赵六……一个一个都叫过来，就和我们把英语、汉语、斯瓦希里语、僮语等语言的句子 1、句子 2、句子 3、句子 4……一个一个搜集起来一样。然后，对召集过来的一个一个的人进行观察研究，归纳总结他们的共同属性，比如说，他们都有两只眼睛、都很高、都讲汉语等，得出来一组他们共有属性集合（27）：

（27）{a，b，c，…，k}

但是，我们无法从这些属性中找出使张三，也使李四、王五，赵六成为一个人的那个或那些属性。但有一点我们是确定无疑的：使张三成为人的东西肯定也是使李四成为人的东西，也是使王五、赵六成为人的东西。果真如此的话，要想知道究竟什么东西使人成为人，就没有必要把所有的人都召集起来对他们进行观察研究，把任何一个人叫来观察研究一下也就可以了，因为任何一个人都具有人的这个种属的区别人以外任何实体的本质属性。所以，我们就可以只把张三叫来就足矣。而后又怎样做才能知道什么东西使张三成为人，即什么东西使任何一个人成为一个人呢？方法就是我们这里说的对比法。按照这种方法，我们可以把张三和任何一个“非人”的实体（比如一张桌子、一条狗）放在一起构成一个对照组：

（28）a．“张三”（张三一词所指称的那个实体）
　　　b．“狗”（狗一词所指称的那个实体）

比较之后，我们便可发现他们之间的最本质差异不在五官、躯体等，而在张三会讲话，那条狗不会讲话。是否会讲话便成了张三与那条狗之间的区别性特征来。实际上也是任何一个人和任何一条狗的种属区别特征。当然，对比不会如此简单，但作为一种方法论讨论也是很说明问题的。一个东西的本质属性往往显露在这个东西与另个东西的最小差异之间。同样道理，我们想知道人类语言的本质，寻找普遍语法的东西也就不必把英语、汉语、斯瓦希里语、僮语中的句子都搜集起来一样，放在一起，在句子和句子之间进行研究，而把句子与非句子组成一个个的最小差异对，进行对比研究会得出更有价值、更深刻的本质的认识。从具体技术上讲，运用这种对比法进行语言研究，就得把一种语言中的一个句子和一个非句子构成一个最小差异对，找出最小的差异点，然后把这些差异点联系起来分析，概括抽象成系统性的规则或原理。这些最终的概括和抽象就可能反映出所有语言的独具的本质属性。西方现代理论

语言学就是这样做的。这就是为什么西方现代理论语言学使用那么多“造得好的句子”(well-formed sentences)、“造得不合法的句子”(ill-formed sentences)、不带星号“*”的句子和带星号“*”的句子、“合乎语法的句子”(grammatical sentences)与“不合乎语法的句子”(ungrammatical sentences)等这些对比性概念的主要原因。人们甚至还可以发现，在西方现代理论语言学文献中还有下面的表示法：

(29) a. what did you buy ?
b. ? what did who buy ?
c. ?? what who bought ?
d. * ? why who bought the book ?
e. * who why bought the book ?

这些句子前面的符号表示句子“造得好坏”的程度(degree of well-formedness)一个问号的要比两个问号的在讲话者的语感中要好一些，两个问号的要比带“*?”的好一些，最不好的，也是根本就不能说的是带一个星号的。因而，在句子(29a)，和非句子(29e)之间出现了(29b，c，d)这样处在两者之间的似句非句的东西，有的西方语言学家称这类句子为“边缘句”(marginal sentences)。这种在讲话者语感中存在着的细微的分别感，他们认为很有理论意义。由于边缘句处在句子与非句子之间，它们的边缘性(marginality)更会显示出句子与非句子间的界限，而这种界限正是寻找普遍语法规则和原理以及语言本质属性所需要的。

对比法的应用范围十分宽阔，任何领域、任何问题都可以使用。在音位学中，在句法学中，在语义学中，在语用学中都得到广泛的应用。在音位学中对比法和最小差异对常常用来鉴别音位学中音素的“区别性特征”(distinctive features)，是描述语言中音位系统构成的最可行的技术方法，在某些语义学中在识别某些语义特征，语义表达解释上的区别性特征也常常使用，在测定、寻找语用因素、语用条件时也是十分有效的方法。

孤立的对比对的研究虽然可以帮助我们认识句子的本质特征，但并不能反映语言系统的全貌，因为语言与句子不是种属与个体的关系，一个句子的存在条件不等于其所属语言的全部，要想把握一种语言的全部本质特征还必须对该语言中所有可能对比对进行研究，从中找出全部句子与非句子的区别性特征。但句子是无限多的，这便引出了方法论上的另一个问题：如何从无限多的事实中找出有限的规律来？

4.2. 事实关联研究

说语言是句子的无限集合只不过是对语言的一种数学描述，不等于说在这个集合中句子是孤立存在的。相反，句子与句子之间有着相互依存的关系，这就是在西方现代理论语言学中常见到的“相关性”(dependency)的一个重要含义。所谓句子的相关性有多种理解，一种是广义的，一种是狭义的。广义的句子相关性指的是在篇章话语中句子与句子之间的某种非

语言的（extralinguistic）事实关系。比如说，句子（1）和句子（2）在下面的篇章中就可能呈现一种以（1）所述事实为因，以（2）所述事实为果的因果关系：

（1）天很冷。
（2）我感冒了。

根据我们的常识信念，听话者和讲话者都会接受这两句话的前因后果的关系，但是这种因果关系不是语言系统本身的规定，而是由语言之外的非语言因素决定的。至于说这种非语言因素究竟有哪些，人们还只能含含混混的说是某些语用的、常识推理的、尚没有成功的形式化的描述。下面两句话的广义相关性就更加含混了：

（3）李先生爱王小姐。
（4）王小姐爱李先生。

假定（3）、（4）都是事实，究竟谁因谁果恐怕谁也说不清楚。对于这种非语言的广义相关性，西方现代理论语言学家中乐观主义者认为非语言相关性最终可以得到形式化的描写，悲观主义者认为很难很难，甚至是不可能的“奥秘”（mysteries）[①]，但他们至少都采取了十分清醒冷静的态度：一方面归纳整理人的推理规则，积极探索这些语用、常识推理的形式化的可能性，另一方面着眼于扩大形式化了的语言系统（主要是句法系统）的句法功能。不管怎么说，西方语言学家都承认语言句法因素与非语言因素的差别。

狭义句子相关性主要表现在一个或一种句子的存在要以另外一个或一种句子的存在为条件。换句话说，一个句子之所以是某种语言中的句子和另外一个句子之所以是这种语言中的句子是相关联的。如果句子 A 不存在，句子 B 也就不会存在了。比如，如果句子（5）不是一个汉语句子，那么句子（6）、（7）和（8）肯定也不会是一句汉语：

（5）有人打跑了那条狗
（6）有人把那条狗打跑了
（7）那条狗被打跑了
（8）那条狗打跑了

所以，研究语言除了研究句子之外，还必须研究句子与句子之间的相互依存关系，即语言事实与语言事实间的依存关系，认为语言不只是句子的自身还是句子间的依存关系。作为一个有机的操作系统，语法不但要对于某种语言中的所有可能的句子作出结构描写，还要抓住句

① “问题”（problem）指人们现在不了解将来会了解的事实、现象等。“奥秘”指的是现在不了解将来永远也不会了解的东西。另一种理解是，有些东西我们现在能“意会”出来，却不能“言传”出来，有些东西我们可以“言传”出来，却不能形式化，有些东西我们从“意会”到“言传”到“形式化”都做得到。我们可以说，永远不能“形式化”的东西为“奥秘”，将来可能“形式化”的东西是“问题”。在研究之前，先估计一下要研究的是“问题”还是“奥秘”，是很有意义的。

间的依存关系。这种句间依存关系在西方现代理论语言学大多数语法系统中都表现为一种"转换"(transformation)关系。比如说，在上面（5—8）这四句话中可以把任何一句话作为"基础"(base)，然后对这个"基础"作出一个结构描写，称作"深层结构"(deep structure)。假如，我们选定（5）为"基础"，它的深层结构如果确定为（9）的样子（只含相关部分，不相关部分略去，并忽略其中对于动词的词法要求）：

(9) NP1　V　NP2

然后，我们在确定（10）这个"转换"规则：

(10) NP1　V　NP2　→　NP1　BA　NP2　V

由于（9）的结构描写符合规则（10）左端输入要求，规则（10）便可把（9）变成规则（10）右端的样子，而（10）的右端又符合（6）的结构描写，最后便可以生成（6）这样的句子。同理，用转换规则（11）可以从（5）或（9）中推出（7），用转换规则（12）可以从（5）推出（8）来：

(11) NP1　V　NP2　→　NP2　BEI　V
(12) NP1　V　NP2　→　NP2　0　V

至于说哪一句话为深层结构，从哪句话转换成哪句话并没有经验性的根据，完全取决于语法系统结构的需要。原则是越简单、生成能力越强越好。这些转换规则的一个作用是反映了这四句话的依存关系。假如（5）不是一个汉语的自然语句，转换规则的左端永远不会有与其相符的输入，也就永远得不到右端的输出，也就不会有（6）、（7）这两句话的存在了。如果（5）、（6）、（7）是一句一句分别生成的，它们之间的依存关系也就不复存在了。不难看出，语法系统中的操作运算反映的是句子之间的依存关系。

句间的相关性不但表现在表面上看来有很多相似之处的句子间，也表现在表面上看来没有多少相似之处，但在系统运算条件上有共同之处的句子结构之间。例如（13）和（14）这两句话之间，初看起来没有什么相似的地方：

(13) John likes himself.
(14) John was arrested.

但是从下面这个推导（14）的过程中，我们却会发现（13）和（14）之间有惊人的相似之处：

(15) i. Somebody arrested John
　　ii. $John_i$ was arrested t_i

(15ii) 是关于（14）这个英文被动句的结构描写式，其中 t 表示主语"John"原来所处的

位置，在这个位置上，它充当主动态动词“arrested”的动作受事者。这个结构描写式是从（15i）推导出来的。人们发现“John”和 t 之间的结构关系和关系条件同句（13）中“John”和“himself”之间的结构关系及关系条件式完全一样的。这一点可以在下面两个最小差异对的对比中得到证实：

(16) a. $John_i$ likes $himself_i$.

b. * $John_i$ thinks she likes $himself_i$.

(17) a. $John_i$ was arrested t_i

b. * $John_i$ thinks she was arrested t_i

(16) 中的对比说明反身代词“himself”只能和它所处的最小子句主语共指，而不能超越它所处的这个子句与主句主语发生任何关系。(17) 中的情景也是这样：t 可以和它所处在的那个最小子句的主语，（17a）中的“John”建立某种联系，但不能越过这个子句与主句主语，(15b) 种的“John”发生任何关系。西方现代理论语言学许多文献把（16)、(17）两句中的反身代词和 t 统称“照应成分”（anaphor），把“照应成分”与其前的相关成分的关系称作“约束”（binding），而照应成分必须受约束。这种“约束”关系只能在“局部的”(local) 结构范围内建立起来。一旦违反了这种局部结构范围的限制，便不会有句子的存在，(16a) 之所以是一个英文句子和（17a）是一个英文的句子的原因一样都是因为它们中的照应成分在局部结构范围内受约束。从照应成分反身代词和 t 本身看，前者在句子中出现的条件和后者在句子中出现的条件是一样的。因此，可以说，如果英文中根本就不存在像（16a）这样的句子，也就不会存在有像（17a）这样的句子。同样道理，如果英文中不存在对应于（16b）这个描写式的句子，也就不会存在对应于（17b）这个描写式的句子。(16)、(17）是相互依存的。不过，这种相互依存的关系和前面的那种相互依存关系不尽相同。比如，像（18）中这两句话间的相互依存关系似乎在句子表面就直接显露了出来：

(18) a. John fixed the car in his garage.

b. It is in his garage that John fixed the car.

语言使用者也好，语言研究者也好，都会凭借自己的语言直觉体会出这两句话相互依存的关系。但对于（16a）和（17a）之间的相互依存关系却不能凭借他们的语言直觉体会出来，得在某种形式化的系统中才能反映出来。我们不妨称前面那种直观的句间关系为“显形相关性”(overt dependency)，后面那种为“隐性相关性”（covert dependency）。在任何语言中，这两种相关性都到处可见。除了句法相关性外，西方现代语言研究中还从语义角度研究句子间的相关性。有一种是人们熟悉的“前设”（presupposition）：

(19) a. What did John buy ?

b. John has bought something.

这两句话相互依存，如果讲话者不认为“John has bought something”就不可能问出“What

did John buy?”这句话来。但讲话者如果认为“John has bought something”，不一定非得问“What did John buy ?”。这种称作“前设”的句间相关性似乎有个方向性，(19a) 以 (19b) 为生存条件，而 (19b) 并不以 (19a) 为生存条件。英语中的情况是这样，汉语亦如此：

(20) a. 张三买了什么?
b. 张三买了些什么。

如果我不认为张三买了些什么东西，我不会问出“张三买了什么?”这句话来。当然，我知道张三买了些什么，我也不一定非问“张三买了什么?”不可。

还有是“蕴涵”(entailment)：

(21) a. John met Mary.
b. John met a woman.

(21a) 蕴涵 (21b)，说“John met Mary”肯定意味着“John met a woman”或“A man met a woman”。但 (21b) 不蕴涵 (21a)，因为说“John met a woman”或说“A man met a woman”不一定意味着“John met Mary”。和前设一样蕴涵也有方向性。

在语言中，句子之间的相关性是西方现代语言研究的一个重要出发点，因而西方语言学家常年花很大力气寻找显性的和隐性的句间联系，尤其是后者，通过对这些相互依存关系条件的概括，期望着把语法系统建立起来。

4.3. 理论逼近法

语言中有那么多句子，句子间又有着那么繁杂的相关性，世界上又有那么多的语言，究竟如何才能建立一个完整的具有普遍语法属性的语言系统，以保证生成所有语言中的所有可能的合乎语法的句子，实现西方现代语言理论所确定的理论目标，把他们所要实现的理论模型建立起来呢?

概括地讲，西方现代语言学家采用一种可称为“逼近法”(approximation) 的理论方法。这个方法的大致内容是这样的：首先对部分语言事实进行观察，从中概括出最简的规则或规则系统，称作第一理论逼近 (the first approximation)，然后再把第一逼近放在另一些相关事实中去检验，看规则是否可行，在事实与第一逼近之间找出规则条件限制上的不充分性（往往是条件限制过松），再对已有规则进行修改，形成为第二理论逼近 (the second approximation)，依此类推，得出第三理论逼近，第四理论逼近，…… 以至最后逼近“真理”。下面举一个具体的例子说明“逼近法”的应用情况和意义。假如我们对下面的事实产生了兴趣：

(1) John likes him.

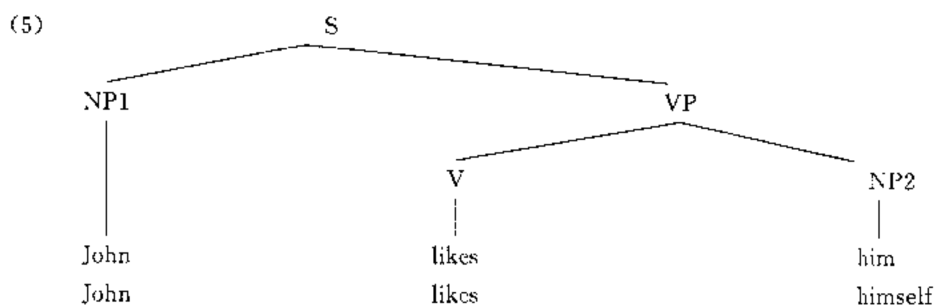
(2) John likes himself.

在这个最小差异对中，两句话在句法结构上没有任何差异，差异只表现在人称代词“him”与反身人称代词“himself”的共称意义（coreferentiality）上，人称代词“him”不能同“John”共指一个人，“himself”只能与“John”共指一个人。这分别是（1）和（2）原来的意义，如果我们写下（1）这些英文字或发出相应与这些文字的声音，而表达“John”和“him”同指一个人的意思，作为声音与意义结合的句子，（1）就不是一句英文，换句话说，我们从不会说出这样的句子。同样，（2）这一行字不能表达“John”和“himself”分别指两个人的意思，就是说，（2）这串声音不会是谈论两个不同的人。为要把（1）和（2）这些文字所承载的意义描写出来，一个常见的方法是使用“共指加标”（coindexation）的形式化构件：

(3) $John_i$ likes him_j

(4) $John_i$ likes $himself_i$

在（3）中，“John”一字上的 i 表示这个词指称 i 这个人，“him”上的 j 表示这个词指称 j 这个人，而 i 在形式上本身就规定它不会等于 j 。因此，（3）这个结构描写式表述了（1）的意义，（4）这个结构描写式表述了（2）的意义，两个结构描写式都是我们想要语法系统给出的输出。想要得出这两个输出的规则至少应含有这样的规定：(A) 反身代词“himself”必须与句子的主语同指，(B) 人称代词不能与句子的主语同指。为了理论逼近描写上的方便，我们先给出相应于（3）和（4）的树形结构式：



用上面结构式中的结构构件，我们可以把想要得到的规定表述成（6），作为第一理论逼近：

(6) 第一理论逼近

a. 反身代词必须被约束。

b. 人称代词必须不被约束。

(当且仅当 X 被 Y 成分统领并同标共指时，X 被约束；否则，X 不被约束。)

有关上述定义中的成分统领的含义见 3.2 节。按照（6）的规定，（5）中的“him”如果带有下标 i，而“John”带有下标 j，作为人称代词的“him”不被约束，与其相应的句子表达

"John" 和 "him" 指的不是一个人，如果 "John" 和 "him" 同标共指，即带有相同的下标时，作为人称代词的 "him" 已被约束，因而违反（6）的规定，得到的是一个不合乎语法的句子，因为表述 "John" 和 "him" 同指的句子是不存在的。同样的道理，（5）中的 "himself" 即受 "John" 成分统领又与它共标同指，因此被约束，符合（6）的规定，与其相应的句子是合乎语法的句子。到此为止，（6）可算作我们的第一理论逼近。这一理论逼近完全可以解释（3）与（4）所代表的事实，同时又可以避免造成（7）和（8）这两个结构描写式所描写事实：

（7） * $John_i$ likes him_i

（8） * $John_i$ likes $himself_j$

现有的第一理论逼近的成功之处在于它一方面给出了（3）、（4）这样句子的原来意义描写又排除了生成（7）、（8）这类非句子的可能，另一方面，还可以解释下面的语言事实：

（9） $John's_j$ $brother_i$ likes him_j

（10） * $John's_i$ $brother_j$ likes him_i

（11） $John's_i$ $brother_j$ likes $himself_j$

（12） * $John's_i$ $brother_j$ likes $himself_i$

原因同（3）、（4）一样工整简单：在（9）中，人称代词 "him" 虽说与 "John" 同标共指但不被成分统领，因而按（6）的规定不被约束。而虽说 "him" 被 "brother" 成分统领但不同标共指，按（6）的规定也就不被约束。所以，相应于（9）的句子是合乎语法的句子。在（10）中，人称代词 "him" 既被 "John's brother" 成份统领又与它同标共指，因而被约束。按（6）的规定，人称代词不能被约束，因而在英语中不会有与（10）相对应的句子。在（11）中，反身代词 "himself" 既被 "John's brother" 成分统领又与它同标共指，所以按（6）的规定被约束，因而在英语中有与其相对应的句子。在（12）中，反身代词 "himself" 虽与 "John" 同标共指，但不受它的成分统领，它虽然被 "John's brother" 成分统领，但它们不同标共指，所以，按（6）的规定，不被约束，在英文中没有与其相对应的句子。作为第一理论逼近的（6）到目前为止是成功的。但是，在下面一些语言事实面前却遇到了麻烦：

（13） * $John_i$ thinks $himself_i$ is intelligent

（14） $John_i$ thinks he_i is intelligent

在（13）中，反身代词 "himself" 即被 "John" 成分统领又与其同标共指，按（6）的规定，它已被约束，在英文中应该有与其相对应的句子，可是，这句话说成（13）的样子不能表达这个意思，因而事实上不存在（13）这样的英文句子。显然，第一理论逼近（6）与（13）所代表的事实不符。（14）中，人称代词 "he" 被 "John" 成分统领又与它同标共指，也就被约束，按（6）的规定人称代词不可以被约束，按理英文中应没有与（14）相对应的句子，可是事实正相反，说（14）这样的话可以表达 "he" 与 "John" 同指一个人的意思。因此，第一理论逼近（6）不但与（13）所代表的事实矛盾，而且与（14）所代表的语言事实矛盾。按理论逼近

法的要求，当第一理论逼近与某些语言事实发生矛盾的时候，不能“另起炉灶”再并列提出另外一套理论规则，而应主动积极地研究矛盾发生的情况，进一步修改已有的理论逼近，必要时，进一步限定理论逼近中的条件，把第一理论逼近提升为第二理论逼近，从而使第二理论逼近不但能概括描述第一理论逼近所覆盖的事实而且能同时能覆盖与第一理论逼近相抵触的新语言事实。按照这种理论逼近的方法论思想，西方理论语言学家仔细研究了（13）、（14）这类事实与（6）之间的矛盾状况，发现（6）中规定过松，反身代词应被约束，人称代词不应该被约束，这似乎没有问题。问题出在它们被约束或不被约束得有个起限定作用的结构域。就（13）、（14）而言，约束反身代词“himself”的是全句主语“John”，含有约束者和被约束者的最小结构是这整个句子，所以说，“himself”被约束的结构域是整个句子。同样，（14）中的人称代词“he”被约束的结构域也是整个句子。反过来讲，如果我们把它们是否应被约束的结构域不看成是整个句子而看成是嵌套在主句中的从句，我们则会看到这样的情景：（一）反身代词“himself”在这个从句的结构域中，即“himself is intelligent”中并没有被约束；（二）人称代词“he”在从句结构域中，即“he is intelligent”中也没有被约束。（一）违反了第一理论逼近（6）关于反身代词应被约束的规定。这就完美地解释了为什么英文中不存在（13）这样句子的事实，同时丝毫没有损伤（6）对原有事实的解释力。在（14）中，人称代词“he”虽说受主句主语的约束，但这个主句主语在“he”所处从句“he is intelligent”之外，因而在从句结构域中，“he”不受约束，作为人称代词“he”没受约束，这本来就符合第一理论逼近的规定，因而，（14）为什么是一句英文就和（13）为什么不是一句英文同时都得到了完美的解释。基于这些考虑，第一理论逼近（6）应改成第二理论逼近（15）：

（15）第二理论逼近

a. 反身代词在其支配范畴内应被约束；
b. 人称代词在其支配范畴内应不被约束。

支配范畴的定义：

X 在下述条件下成为 Y 的支配范畴：

X 是含有 Y 的最小句子。

从第一理论逼近到第二理论逼近，理论所覆盖的事实范围扩大了，理论解释上的充分性保持不变，都是对人称代词和反身代词句法分布（syntactic distribution）的限制，而理论描写上的充分性增强了。但是，第二理论逼近仍不是终极的真理，在下面的事实面前又遇到了麻烦：

(16) $Mary_j$ was upset by $John's_i$ criticism of $himself_i$

(17) $Mary_i$ was upset by $John's_j$ criticism of her_i

在（16）这个结构描写式中，反身代词“himself”和“John”同标共指，并不和它所处

的最小句子，即整个句子中的主语“Mary”同标共指，也就是说，在它的支配范畴内不被约束，这样便违反了第二理论逼近中关于反身代词在其支配范畴内应被约束的规定。然而，（16）却是一句合乎英文句法的句子。第二理论逼近和（16）这个新的事实发生了矛盾。在（17）中，人称代词“her”在其支配范畴内，即整个句子中，被主语“Mary”约束，按第二理论逼近关于人称代词在其支配范畴内应不被约束的规定，（17）不应是一个英文句子。可是，（17）偏偏是一句造得很好的句子，这个事实又构成了对于第二理论逼近中关于人称代词在其支配范畴内应不被约束这一规定的挑战。这就逼迫我们得根据（16）、（17）所代表的新事实、新情况对第二理论逼近作进一步修改。有两种修改的可能方法：一是修改关于反身代词和人称代词是否应被约束的规定，一是进一步限定它们是否应被约束的结构域，即支配范畴的定义。如果采纳第一种办法，就要重打鼓另开张，就得从根本上改动第一和第二理论逼近的核心内容，这就必然会丧失第一和第二理论逼近已获得的解释上的充分性和描写上的充分性。这显然要花很大的理论代价，是不可取的。人们不由地都采用了第二种逼近的办法，就是在不改动关于人称代词和反身代词是否应被约束的要求的基础上，修改受约束或不受约束的结构域，即支配范畴。修改后我们便得到了（18）这个第三理论逼近：

（18）第三理论逼近

a. 反身代词在其支配范畴内应受约束，
b. 人称代词在其支配范畴内应不受约束

支配范畴的定义：

X 在下述条件下成为 Y 的支配范畴：

X 是含有以下三个成分的最小结构

i. Y
ii. 支配成分
iii. SUBJECT①

其中，支配成分为：I，V，P；
SUBJECT 为：i. 句子的主语（subject）
ii. AGR
iii. NP of NP

按照现在这个第三理论逼近的规定，在（16）中，反身代词“himself”的支配范畴是它所处在的名词短语结构“John's criticism of himself”，而不是整个句子，因为这个名词短语结构是含有 Y（himself），支配成分（“of”）和 SUBJECT（John's）的最小结构，因此是它的

① SUBJECT 包括表示“一致性”的 INFL / AGR 及名词词组中结构位置最高的名词性成分。

支配范畴，而在这个支配范畴内，“himself”即被“John”成分统领又和“John”同标共指，因而也就在这个支配范畴内受约束。在（17）中，人称代词“her”的支配范畴是它所处在的那个名词短语结构“John's criticism of her”，因为这个名词短语结构是含有Y（her），支配成分（of）和SUBJECT（John）的最小结构，而在这个支配范畴内“her”与“John”不同标共指，也就不受约束。这两例子都符合第三理论逼近关于人称代词和反身代词的规定，事实上也是两个造得好的英文句子。这样，第三理论逼近不但保持了第一、第二理论逼近原有的理论解释力和事实覆盖力，而且又进一步解释了第一、第二理论逼近不能覆盖的新事实。实际上，第三逼近就是在西方现代理论语言学界极为有影响的“约束论”（The Binding Theory）的雏形。第三理论逼近仍旧不是理论的终点。目前，“约束论”还在不断地在新的语言事实面前得到修改，不但在一种语言中的新事实面前还要在多种语言的新事实面前得到修改。

比如说，人们发现第三理论逼近，即“约束论”不能完全充分地解释汉语中有关反身代词“自己”和“他/我/她/我们/他们/她们－自己”某些现象。以下面一句话为例：

（19）张三认为李四不喜欢自己

在这句话里，我们的汉语语言直觉告诉我们：“自己”既可以指“张三”又可以指“李四”。当“自己”指“张三”时，它的约束支配范畴为整个句子，可当它指“李四”，它的约束支配范畴是以“李四”为主语的从句。按“约束论”的规定，如果“自己”是反身代词，那么它的约束支配范畴只能是那个从句，因为这个从句是最小的含有“自己”本身，支配成分（动词“喜欢”）和SUBJECT的成分，而没有可能受主句主语“张三”的约束。相反，如果把汉语的“自己”看成是人称代词，按照“约束论”人称代词在其支配范畴内应不受约束的规定，它则没有可能与从句主语“李四”同标共指，因为它的支配范畴不可能再小于这个从句了。总之，无论是把“自己”看成反身代词，还是人称代词，都不能在“约束论”中得到圆满的解释。日语、鲜语等语言中也出现了类似的现象。是“约束论”本身的理论解释力不够强呢，还是没有能充分地利用“约束论”中已有的理论解释力呢？从这两方面考虑问题的都有。认为“约束论”本身理论解释力不够的提出了“概括式约束论”（Generalized Binding Theory）；认为现有“约束论”的理论解释力仍可以进一步挖掘的提出关于反身代词随INFL抽象移动的看法。另外有人认为这两种办法都不能补救“约束论”的不足，提出了关于汉语、鲜语等语言中的反身代词本身有不同于英文反身代词的属性的看法，认为汉语、鲜语中的反身代词具有人称代词和反身代词的双重属性，当用作人称代词时在句法中的表现和英文中的人称代词一样，在它们的支配范畴内不受约束，当用作反身代词时和英文中的反身代词一样应在它们的支配范畴内受约束。这种属性称作“logophoricity”这似乎能解决像（19）这类句子提出的问题。正因为存在着这些悬而未决的问题，需要我们把第三理论逼近推向第四、第五逼近。

从上面关于理论逼近的论述中我们可以看出，西方现代理论语言学在对待“例外”（exception）和“反例”（counterexample）的问题上采取了与传统语言理论不一样的态度。西方现代理论十分欢迎“例外”和“反例”，认为与已有规则相矛盾的“例外”和“反例”是推动理论发展的动力，是进行理论逼近的依据，从不采取为维护已有理论规则把“例外”和“反例”拒之门外，而是主动积极寻找“例外”，把已有理论放在“例外”和“反例”面前，接纳它们，

研究它们，从中找出理论的不足，长期不懈地在下面这种理论建立模式中追求理论上的更大完美：

（20）立论 — 例外/反例 n → 理论 n — 例外/反例 n+1 → 理论 n+1

4.4. 在语言直觉知识面前

和传统语言理论相比，西方现代理论语言学理论方法的具体特征还鲜明地表现在对待人的语言直觉知识上。可以说，在对待语言直觉的态度上传统语言理论和包括描写性语言理论在内的西方现代语言理论是根本不同的。传统语言学家，尤其是“教学语法学家”(pedagogic grammarians)、“规定主义语法学家”(prescriptive grammarians)和结构主义语言学对待语言直觉问题一般所采取的态度是这样的：他们凭借它们本人的某种具体语言的直觉知识，认识和分析语言事实，然后把认识分析的结果奉献给使用者（如外语教师、外语学生等），这些使用者同样凭借着它们已有的语言直觉接受和认识语言研究者凭借它们的语言直觉所得出的研究结果。比如说，当研究的语言是本族语时，语言研究者会凭借着自己关于这个本族语的直觉知识把一个句子切分成“主语部分”和“谓语部分”，然后把这个研究结果告诉一个本族语的使用者（学生），让它们分析该语言中的一个句子，这个学生便会凭借着他的本族语关于“什么是主语部分”和“什么是谓语部分”的直觉，给出一个语言研究者所期待的结果。同样，语言研究者会凭借自己的本族语直觉发现语言中有（1）这类歧义的句子：

（1）鸡吃完了。

从而提出了“歧义句”理论概念，写在语言理论中。而后语言理论使用者便会使用这个概念让学生分析下面的句子是否歧义：

（2）自行车没锁。
（3）自行车没锁好。
（4）自行车没锁更好。
（5）自行车锁没锁？
（6）自行车有锁。
（7）自行车锁好了。
（8）自行车没锁了。
（9）自行车没了锁。
（10）自行车锁了。
（11）自行车锁好了。
（12）自行车锁没了。
（13）自行车好锁。

凭借着自己的语言直觉，学生会判断出（2）、（3）、（4）、（7）和（11）都是歧义句，（5）、（6）、（8）、（9）、（10）、（12）和（13）都不是歧义句。假如一个学生根本没有汉语语言直觉（例如一个学习汉语但还没有获得汉语直觉知识的外国学生）无法接受"歧义"这个理论概念，也无法对这些句子作出是否歧义判断。在本族语研究中是这样，在"教学语法"的外语研究中也是如此。这种凭借语言直觉研究语言的方法没有任何不当之处，实际上，任何方法的语言研究（甚至是任何一种科学研究）无一例外地都离不开语言直觉。可是，有一个十分值得考虑的问题是，要不要揭示语言直觉的内容，要不要回答为什么语言研究者和语言学习者都有同样的语言直觉，要不要回答语言研究者和语言学习者凭借着什么样的语言直觉断定每一句话无一例外的都能分成主语部分和谓语部分，都能辨认出歧义句来。在数学、物理、生物这些学科里，似乎没有必要回答什么是直觉知识的问题，没有必要考虑为什么人人都觉得一和一相加是二，人人都会觉得天平一端上的砝码高于天平另一端上的砝码，人人都会觉得鸟会飞、鱼会游水，而不需要在数学理论中写下关于"一"、"二"、"加"的直觉感受，也不需要在物理学中回答人是怎样觉得天平一头高一头低的，也不需要在生物学中谈及人是怎样觉得鸟在飞的，可是，在语言研究中，这些都是在立论之前必须回答的方法论问题。这是因为语言研究和其他学科研究的一个根本不同之处是：语言研究中的研究工具和研究对象都是同一个东西——语言，而其他学科的研究工具和研究对象有着天然的客观距离。由于传统语言立论的研究对象是语言事实，不是语言事实的成因，因而也就不顾及、不研究观察语言事实所凭借的语言直觉，就像生物学不顾及、不研究鸟儿飞翔所凭借的视觉直觉一样。而西方现代理论语言学不是这样。他们不但（也只能这样）凭借语言直觉，而且研究语言直觉，揭示语言直觉的内容，为语言直觉给出形式化的系统描写，甚至回答语言直觉是怎样获得的问题。他们建立的理论就是关于语言直觉的理论，它们所建立的语法系统就是关于语言直觉的系统，它们所提出的规则就是关于语言直觉的规则。下面对西方现代理论语言学所涉及的一些语言直觉知识内容及如何描述的做一个简单的归纳。

第一种语言直觉知识内容是句子的"语法性"。这在前面已经提到过，这里只重复一个例子：

（14）　张三睡了。
（15）＊张三睡了李四。

对于这种直觉知识，一般用"选择限制"来描述。在语法系统中，按这种限制动词"睡"只能出现在"其后没有名词短语"的语境环境中，不能出现在"有名词短语"的语境中。正是人的语言直觉系统中有这种"限制"，所以语言研究者和语言是用者都会有关于（14）是汉语，（15）不是汉语的感觉。

第二种语言直觉知识是刚才提到的关于歧义的辨认。这种对于歧义辨认的直觉知识能力表现多种多样。最常见的是人们所熟悉的词汇歧义（lexical ambiguity）、结构歧义（syntactic ambiguity）。（16）、（17）是两个例子：

（16）打a. 打人、狗

b. 打油、车票
c. 打球、麻将
d. 打家具
e. 打瞌睡
f. 打井、洞
g. 打毛衣
h. 打官司、离婚
i. 打麦子
j. 打更、工
k. 打前站
l. 打住
m. 打夯、桩
n. 打江山
o. 打气（给人打气）
p. 打鱼
q. 打针
r. 打（了一张）牌
s. 打气（给车打气）

(17) a. [张三和李四] 的 老师
a'. [张三] 和 [李四的老师]
b. 这只猫不在床下捉老鼠。

关于（17—b）这句话的否定词“不”的“否定域”至少有下面几种释义性语义解释：

(i) 那只猫在床下捉老鼠。
(ii) 这只猫在床上捉老鼠。
(iii) 这只猫在床下捉苍蝇。
(iv) 这只猫在床下玩皮球。
(v) 这只猫在洞里捉老鼠。
(vi) 这只鹰在床下捉老鼠。
(vii) 这条狗在床下捉老鼠。
(viii) 这只猫在床下吃老鼠。
(ix) 这只猫在柜下捉老鼠。

在词汇歧义的问题上，西方语言学家一般认为，在语言直觉知识系统中，“一词一义”不存在所谓的“一词多义”现象。换句话说，在人脑的那个词库里没有相应于常见一般词典中“打”这个词典意义上的词，而只有“打人”的那个“打”，“打油”的那个“打”，“打官司”的那个“打”等。一个意义和另外一个意义碰巧表达为一个声音文字符号，造成了歧义。基

于这种认识，一般词典意义上的多义词都分别加上一个下标，表示不同的意义。如，“打 1”表示“打人”的“打”，“打 2”表示“打油”的“打”等。当一个词进入一个句子结构描写中时就带有表示特定意义的标号，使每个结构描写式都对应于一个非歧义的句子。也有人认为，这些表示不同的意义的特定标号是在句法环境中加上的。当多义词“打”进入一句话后，看它所处的语境环境，如果其后邻的词是“人”，这个“打”字就表示一个意义，如果后邻的词是“鱼”，那么这个“打”字就表示另外一个意思。不过这都不是语言真实心理发生过程的描写，而只是一种理论上的解释。一句话在讲话者的脑子里永远不是歧义的。歧义只发生在听话者方面。听话人分辨歧义句，的确是人的语言直觉知识的一个方面。关于这个问题，语言研究者只能作出笼统的回答：靠语境。可是，至于究竟是怎样靠语境，语境中有哪些相关因素等问题，离形式化描写还有相当大的一段距离，甚至有人说是不可能的。

结构歧义在西方现代理论语言学中是十分重要的研究内容。这种歧义现象涉及到语言的各个方面。举几个典型的例子：

(18) 张三说李四不爱惜自己。

这句话既可包含“张三说李四不爱惜李四他自己”的意思，又可包含“张三说李四不爱惜张三他自己”的意思，不管是从讲话者的角度看还是从听话者的角度看，这个歧义的句子实际上是两个不歧义的句子，其结构描写式分别为 (19) 和 (20)：

(19) 张三$_i$ 说 李四$_j$不爱惜 自己$_j$

(20) 张三$_i$ 说 李四$_j$不爱惜 自己$_i$

再看 (21)：

(21) 他不喜欢什么

这句话也可能有两个不同的含义：一是说“他什么也不喜欢”，一是问“他不喜欢的东西是什么”。同样，不管是从讲话者的角度看还是从听话者的角度看，这个在语言研究者看来是歧义的句子实际上是两个表达不同意义的句子，可分别在语法系统中描写成 (22) 和 (23)：

(22) [[什么]$_i$ [他不喜欢 t_i]]

(23) [Op_i] [他不喜欢 什么$_i$]]

忽略这两个结构描写式中的技术细节和各符号意义不谈，(22) 表述的是一个特殊疑问句，(23) 表述的是“什么”这个词不含有疑问词的特征，而只含有“逻辑量词”(quantifier) 的特征，表述“任何东西”的意思。

总之，西方现代语言学想把人们对语言的直觉知识和能力首先“意会”出来，再用自然语言“言传”出来，再尽可能地“形式化”。

4.5. 比较研究法

西方现代理论语言学所刻意追求的不只是为一种具体语言提供一个形式的语法操作系统，还要为普遍语法提供一个理论模型，从中能够推演出所有可能人类具体语言的具体语法。这个普遍语法的理论模型包括为获得任何一种可能人类语言所遵循的有限的普遍语法原理和参数，一方面对什么是可能的人类语言限制具有充分的解释力，一方面又保证足以覆盖所有具体语言的特异性。

虽然说，大多数西方语言学家们都认为普遍语法的东西可以在一种具体语言中全部找到，这就和我们可以在任何一个人身上找到人的普遍属性一样。共性确实寓于个性之中。但是，从方法论的角度上讲，很难在一个孤立的个体中发现共性的东西，很难从一种具体语言中辨别出所有人类语言所共有的东西，也很难从人获得一种具体语言能力所遵循的东西中辨别出获得任何一种具体语言能力所遵循的东西，虽然前者（个性）是后者（共性）的体现。这样，寻找普遍语法的东西就不可避免地在多种具体语言的现实中寻找。"跨语研究"（cross-linguistic study）或"比较研究"（comparative study）就成为西方现代理论语言学另一个最基本的研究方法。

从总的比较研究方向上讲，比较研究大体可分成两大类：一类是现象事实上的归纳概括上的比较，具有明显的描写性，简称事实比较（factual comparison）或事实比附；另一类是关于系统理论普遍语法原理在具体语言中的可行性的比较，简称理论性比较。先看第一类的比较。在这类比较研究中可见下几种比较的内容和方法。

事实性比较内容在两种语言间都是显性的、有标记的、结构上的异同。比较语序就是一个最典型的例子，英语、汉语等语言共有下面一些句法结构：

(1) 主 — 谓 — 宾
　　a. [John] [fixed] [the car].
　　b. [张三] [修好了] [那部车]。

(2) 名词短语的内部结构：（限定词）—（数量词）—（形容词）—（名词）
　　a. [those] [three] [cute] [babies]
　　b. [那] [三个] [乖] [孩子]

(3)（助动词）—（动词）
　　a. [can] [fly]
　　b. [会] [飞]

而在下面一些结构上英语和汉语表现出不同：

(4) 英语中的定语从句在名词中心词右侧，汉语定语从句在其所修饰的名词中心词左侧：

a. [the baby] [everyone loves]
b. [人人都喜欢的] [那个孩子]

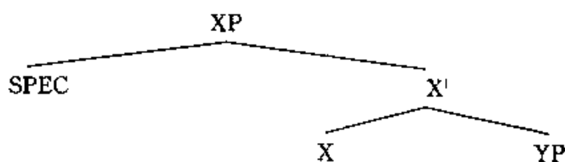
(5) 英语状语成分可出现在谓语动词短语的右侧，汉语状语成分则出现在谓语动词短语的左侧：
a. We [had our dinner] [in the restaurant]
b. 我们 [在那家餐馆] [吃的晚饭]

这样，在上面的比较中，我们一方面看到了英汉两种语言结构上的相同处，同时也比较出了英汉两种语言间结构上的不同处。按照这个方向比较下去，就可以找到英汉两种语言在句子结构上的基本异同，到此就列举出了跨语可比的事实，然后尽可能地对列举出来的可比事实归纳概括。归纳概括的程度越高越好，而做到这一点，用形式化的办法往往显得简洁方便。比如说，从上面列举的可比事实中，可以归纳概括成以下几点：

(6) 英语和汉语的补语 (complement) 都在其相关中心词的右侧，而英语中的“嫁接成分”(adjunct) 在右，汉语中的“嫁接成分”在左。

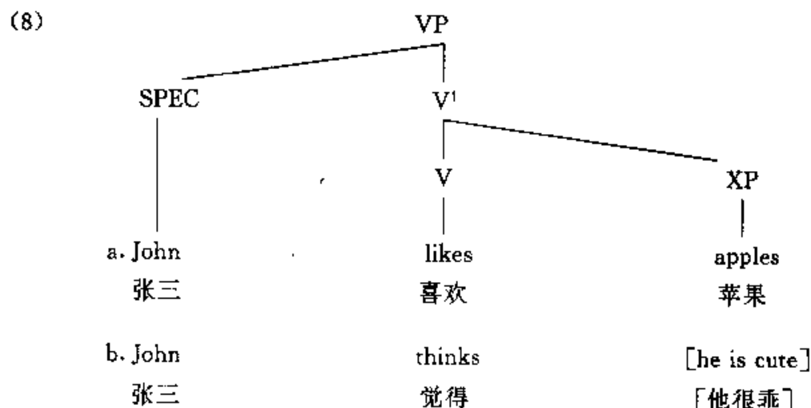
这个归纳概括用形式化的方法可以表达成下面的样子：

(7) 汉语和英语



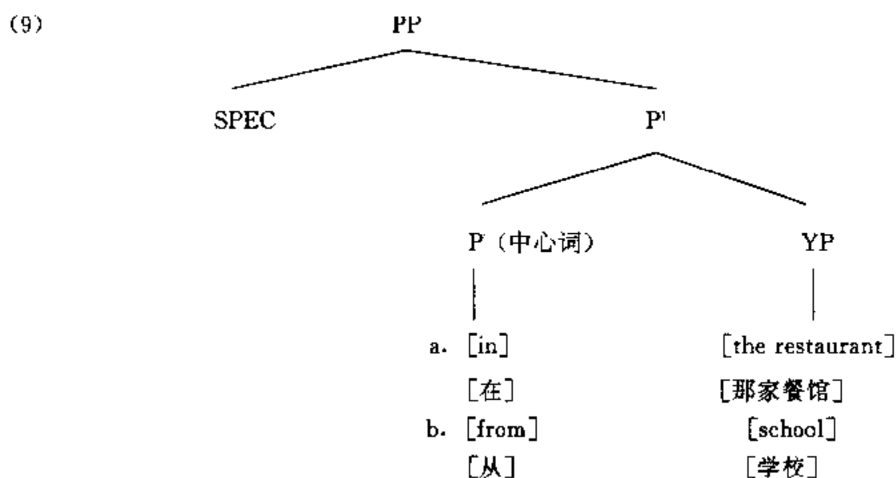
其中，X 可为 V、AUX、P 等。当 X=V 时，可以看到下面的情况：

(8)



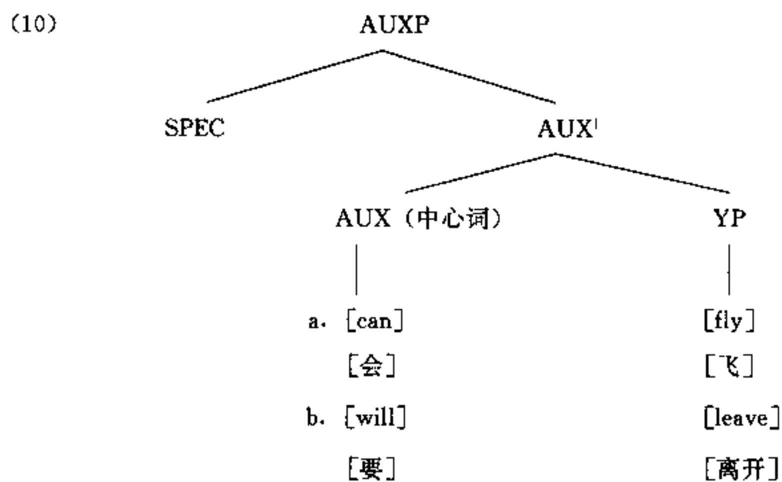
当 X=PP 时，可以看到下面的情况：

(9)



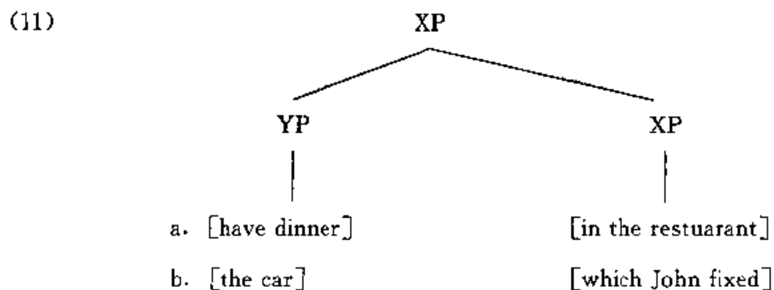
当 X=AUX 时，两种语言出现下面的情景：

(10)

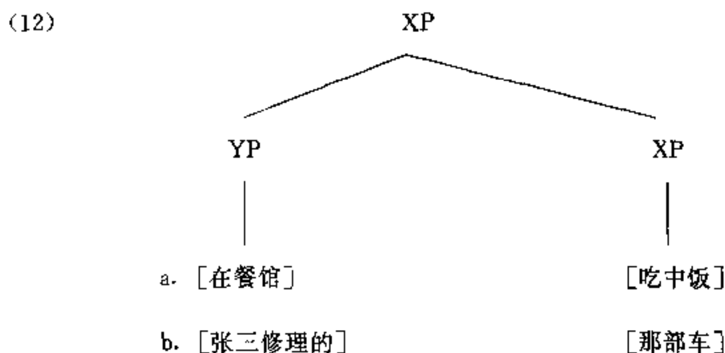


英汉语言在"嫁接成分"位置上所表现出的不同可概括地描写成下面的样子：

(11)



(12)



在(10—12)的基础上，我们可规定(13)为关于语言语序的第一理论逼近。

(13) a. 中心词在左

b. 嫁接位置为最大映射

然后再把这个第一理论逼近放到其他语言中，再得出关于语序的第二个理论逼近。假如，这种语言是日语，我们要做的不是在英语、汉语和日语之间进行比较而是在第一理论逼近和日语的基本语序事实之间加以比较。

再有一种常见的跨语比较研究的情况是在一种语言中为有标记(marked)的事实与另一种语言中无标记的(unmarked)事实之间进行比较。所谓有标记的是指那些有显性形态特征的词汇成分或句法成分，比如英文中主语要和谓语动词在人称、数上保持一致(agreement)，这种一致性有显性的形态标志，例如：

(14) a. He likes Mary.

b. They like Mary.

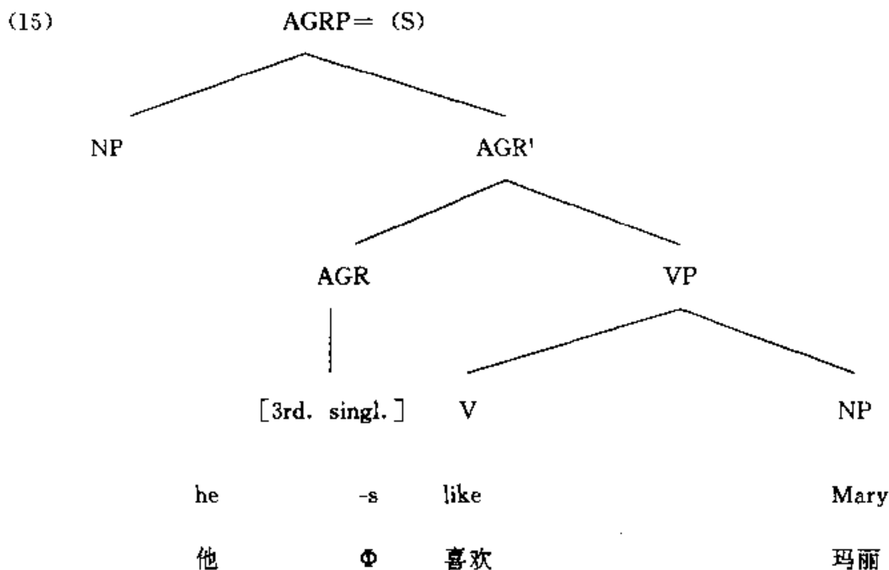
而在汉语动词没有这种人称、数的形态变化：

(14) c. 他喜欢李小姐

d. 他们喜欢李小姐

还比如，英文中有分别表示"特指"(definiteness)和"泛指"(indefiniteness)的不同词汇形态："the"和"a/an"，而汉语中没有严格意义上的对应词汇。这种一种语言中有，另一种语言中无的现象，在传统语法中有详细的描写。但在解释这种"你有我无"的跨语现象时，西方现代语言理论与传统语言理论有很大的不同。一般说来，传统语法认为"有便是有，无便是无"，这是两种语言的特异处，没有共通的地方可言。这的确是千真万确的事实。在西方现代理论语言学中，有者为有，无者也为有，差别不过是前者有在"明处"，后者有在"暗处"，"暗处"无标记。无标记的有一般都冠以"隐性的"或"抽象的"(abstract)，对有标记的称之为"显性的"。在句法系统中，一种语言中隐性的成分和另一种语言中显性的成分一样在结构描写式中都有同样的表达。比如，汉语中隐性的"人称一致"就和英文中显性的"人称一致"都被描写为 AGR 这一结构成分：

(15)



这种以一种语言中显性成分为基准来断定另一种语言中有其相应的隐性成分并给予同样的结构描写的理论出发点是，不管隐性的也好显性的也好，同它们相对应的范畴概念都是一样的，假如这种范畴概念是不可定义的，就有理由把他们都看成是一种人类认识上的初始元，因而也是使用各种语言的人所共有的，可以看作是语法形式系统的初始元素。所以，西方现代理论语言学多半都把表示“一致性”的成分看成是普遍语法的东西，不过在英语、法语、西班牙语等语言中是有相应形式标记的，而在汉语、日语、朝鲜语等语言中是没有相应的形式标记的。两者都是“一致性成分”AGR 的不同变体。也有学者反对这种看法，认为不能因为一种语言中有标记的成分就一定会在另一种语言中以无标记的形式存在，而正好相反，有者只能是有形态的，无形态者则为无。这两者一有一无正是两种语言的差异。如果系统地多方面地比较两种语言间的显性/隐性成份的差异，我们会发现这些差异往往表现在“虚词”类或“功能”类（functional elements）成分上，“实词”部分几乎看不到这种情景。这似乎是自然的看法。(这里讲两种语言间实词中不会你有他无，指的是语言成分的范畴类，而不是一个两个单词。英语中本没有相应于“阴、阳”的词语，汉语中本没有相应于英语中“God”的词语。实词是个开放性的无限集合，而虚词成分是个封闭性的有限集合。在无限的范围内谈“有”没有意义，而在有限的范围内才会有“有无”可言。笔者认为，后一种意见更有道理。

还有一种极为常见的比较。这种比较的内容多半是语义上的比较，其中还可以细分成为两小类：一类是语义关系上共性的比较，一类是语义解释（semantic interpretation）上共性的比较。前者见诸于实义动词的选择限定上或“概念结构”（conceptual structure，argument structure）上。请看下面一些动词在英汉两种语言间比较的例子：

(16)

英语	汉语	可跟随成分	不可跟随成分
die	死	0	NP/S
kick	踢	NP	S/AP

believe	相信	S	WH-S
put	放	NP PP	NP
wonder	想知道	WH-S	S/NP
give	给	NP NP	NP/S

违反这些规定，英语不成立的句子在汉语中相应也不成立，在英语中成立的在汉语中也成立：

(17) a. John died.
b. 张三死了。

c. * John died [NP Mary].
d. * 张三死了 [NP 李四]。

e. * John died [S Mary survived].
f. * 张三死了 [S 李四活了]

(18) a. John kicked [NP the ball].
b. 张三在踢 [NP 球]。

c. * John kicked [S Mary was pretty].
d. * 张三在踢 [S 李小姐很漂亮]。

e. * John kicked [AP intelligent].
f. * 张三在踢 [AP 聪明]。

(19) a. John believed [S Mary was pretty].
b. 张三相信 [S 李小姐很漂亮]。
c. * John believed [WH-S where Mary had her dinner].
d. * 张三相信 [WH-S 李小姐在哪里吃过饭]。

(20) a. John put [NP the letter] [PP on the table].
b. 张三放 [PP 在桌子上] [NP 一封信]。
c. * John put [NP the letter].
d. * 张三放 [NP 一封信]。

(21) a. John gave [NP me] [NP a new car].
b. 张三给了 [NP 我] [NP 一部新车].
c. * John gave [NP me].

d. * 张三给了 [NP 我]。

从这些例子可以看出来，英汉两种语言间这些实义动词在选择什么成分可跟在后面的问题上表现出完全一样的属性。许多语言学家在比较、研究了世界上许多语言之后发现，意义一样的实词间的句法属性没有差别。

第二种语义解释上的比较多见于有语义域关系的句子间。下面是一些典型的例子：

(22) a. Who does John like?
b. Who does Mary think John likes?

(23) a. 张三喜欢谁?
b. 李小姐认为张三喜欢谁?

(24) a. John promised Mary to start earlier.
b. John persuaded Mary to start earlier.

(25) a. 张三答应李小姐早动身。
b. 张三劝李小姐早动身。

在 (22) 与 (23) 之间的比较，可以发现英语疑问词“who”和汉语疑问词“谁”在句子中的位置虽然不同，前者远离与其相关的动词放在句首，后者紧邻与其相关的动词在原地不动，但它们在句子中的语义域 (semantic scope) 是完全一样的：在英文中，问“Who does Mary think John likes?”意思是说：“把 Mary 认为 John 喜欢的人告诉我，即使 John 喜欢什么人，但 Mary 不认为 John 喜欢他，我也不想知道。这就和问“Who does John like? ”不一样，问“Who does John like?”，就不一定得有“Mary 认为的”的这个限定。比较之下，汉语在 (23) 中情景也是这样。问“张三喜欢谁?”，你随便说好了，即使你告诉我的是一个李小姐认为张三不喜欢的人，我也不会感到奇怪。但如果我问的是“李小姐认为张三喜欢谁?”，你就不能把李小姐认为张三不喜欢的人告诉我，因为我期待着你告诉我张三喜欢的那个人要限制在李小姐认为是如此的范围内。这样，当语法系统要给出关于 (22b) 和 (23b) 这两句问话的语义结构描写式时，就得找出一种统一的形式，即属于普遍语法的东西。这种表达语义的统一形式，在生成语法等语言理论中称作“逻辑式” (Logical Form)，在其他一些语言理论中称作“语义解释” (Semantic Interpretation 或 Semantic Construal)。在 (24) 与 (25) 的比较中，可以看到英文中的“promise”和“persuade”同汉语中的“答应”和“劝”一样引出不同的语义解释：含有“promise”的句子和含有“答应”的句子一样——“谁答应，谁早动身”；含有“persuade”的句子和含有“劝”的句子一样——“劝谁，谁早动身”。既然两种语言共有一种语义解释，就应在语法系统中得到统一形式描写。这种统一的描写式，在西方现代理论语言学中大多数都用特定的“空范畴”标示符 PRO 加下标的方式来表达：

(26) a. $John_i$ promised $Mary_j$ PRO_i to start earlier.

b. $John_i$ persuaded $Mary_j$ PRO_j to start earlier.

(27) a. 张三$_i$答应 李小姐$_j$ PRO_i早动身。

b. 张三$_i$劝 李小姐$_j$ PRO_j早动身。

这些跨语比较并不是在一个个孤立的语言事实间进行的，而是放在一个系统进行系统的比较。这表现在以下几个方面。第一，比较的结果都必须表达成形式化的东西，必须用系统中已有的结构形式构件表达。比如说，当语法系统已有PRO这个形式构件后，就可以使用这个形式构件把（24）、（25）描写成（26）、（27）的样子。这种比较研究方向具有演绎的性质，即从系统公理或原理出发，把已有普遍语法原理放在未及语言中检验其真伪或真伪程度。这在西方现代理论语言学中，尤其在生成语法中是比较研究的主要内容，而且从中证实了许多共性极强的普遍语法原理。最有名的有生成语法中的“比邻原理”（the Subjacency Principle）、“映射原理”（the Projection Principle）、“外移域条件”（the Conditions on Extraction Domain）、“空范畴原理”（the Empty Category Principle）、“约束论”（the Binding Theory）等，概括性短语语法中的“中心词特征规则”（Head Feature Convention）、“根基特征原理”（Root Feature Principle）、“控制匹配原理”（Control Agreement Principle）等，词项功能语法中的“独具性条件”（Uniqueness Condition）、“完全性条件”（Completeness Condition）和“一致性条件”（Coherence Condition）等。

如果语法系统中没有可用的形式构件，则可以依据比较的结果在形式系统中增加新的结构形式构件。这种比较研究方向具有归纳的性质，即从语言的具体事实出发，把比较的结果归纳概括成可能具有普遍意义的原理，放在系统中去，成为系统的一个有机部分。比如说，根据（22）、（23）英汉特殊疑问句的比较结果，人们就想出了一种叫逻辑式的结构描写式并把这种描写式加进了语法系统，用来对语言中共有的特殊疑问词的语义域进行解释描述。这就在语法系统中出现了一个为人们所熟知的表达层次——逻辑式。归纳性的比较和演绎性的比较在西方现代理论语言学中兼顾并重。不同理论发展时期，不同的研究兴趣会有所侧重，或取归纳性的比较，或取演绎性的比较。一般说来，在目前西方现代语言学文献中演绎性的比较研究为多，这和演绎性比较研究同理论逼近方法有着密切的联系有关。这里讲的都是想比较出来个普遍性的情况。当比较的结果出现差异时，尽量把差异看成是一个普遍语法共项的变体，而不轻易地宣布那是语言的特异部分，从而提高语法系统的普遍语法性，在普遍语法原理所及范围内，把语言特异部分尽可能地参数化。

跨语比较研究自然会导至多种语言间的相互比较，这就引出了西方现代理论语言学的多语研究的特点。这一点在生成语法研究中尤为明显。格语法、概括性短语结构语法、功能语法、关系语法和词项功能语法等语言理论目前都局限英语这一种语言中，没见有其他语言参与理论建立。生成语法研究所涉及的语言已有近百种之多，直接参与生成语法理论建立的，对生成语法理论系统建立有重大贡献的语言包括英语、汉语、法语，西班牙语、日语、朝鲜语、挪威语、沃尔波里语（Warlpiri）、荷兰语、纳瓦霍（Navajo）、意大利语、德语等。

比较研究的理论意义在于：（一）可以增强普遍语法原理的普遍性，进而增强语言理论的理论解释力；（二）可以发现在一种语言不易发现的具有普遍语法意义的东西，进而丰富普遍语法的理论内容；（三）可以更多地找出语言的特异处，使普遍原理的参数化更加概括和优化，

增强理论描写力，扩大理论覆盖范围；（四）促进语言类型学（Typology）研究。

4.6. 二元属性描写法

西方现代理论语言学还有一个被广泛应用具体的研究方法，可概括地称为“二元属性描写法”(Binary Feature System)，简称“二元法”。这种研究方法的基本要领是把系统中的表示属性理论概念都看成是一个概念的两个对立的互相排斥的值，一个为正，一个为负，在正负之间没有任何其他属性。先从一个常识中的例子谈起。

我们说某一件东西X“长”，另一件东西Y“短”。在这里我们使用了“长”和“短”两个概念。像（1）中那样我们画一条线，线上有许多点，线上最左端的那个点表示“长”，最右端的那个点表示“短”：

（1）长　　　　　　　　　　　　　　　　　　　　　　　　　　短

•—•—•—•—•—•—•—•—•—•—•—•—•—•—•—•

a　b　c　d　e　f　g　h　i　j　k　l　m　n　o　p

可以想像，这条线上的点可以是无穷无尽的，在a、b之间还可以列出许许多多的点来，在l、m之间，在k、l之间也可以列出许许多多点来。假定i为这条线的中心点，人们自然会认为i点以左的所有点多可代表“长”，i点以右的都可以称作为“短”，而i点可以说是“不长不短”。可是h和i之间的及i和j之间的呢？由于在这两点之间的点是无穷无尽的，“长”和“短”的界限究竟在哪里呢？模糊数学和所谓的“模糊语言学”使用了模糊集合的概念对这个“长”“短”之间的模糊界限作了说明，这可能适合以模糊集合演算为主的理论体系，但不适合用于明晰的形式化语法系统。另外，在使用两个表面上看起来是对立的但界限不清的概念描写事物属性时，实际上这两个概念是并列的，因此也允许在这两个概念之间出现界于两者之间的概念。比如，说X“长”，说Y“短”，虽然不能再说Z“又长又短”，但不排除可以说Z“不长不短”。这样，我们就不能使用“长”和“短”这两个概念穷尽所有的事物，因为在“长”“短”之间还有“不长不短”。因为在“长”“短”之间还有“不长不短”。同样道理，我们也不能使用“好”“坏”这两个概念穷尽所有的事物，在“好”“坏”之间还有“不好不坏”。我们可以把这种使用两个或多个并列概念来描写事物属性的方法叫做“多元属性描写法”。这种“多元属性描写法”的特点在下面的例子中表现得更为明显了。

（2）深黑　黑　浅黑　　深灰　灰　浅灰　　乳白　　　白

—•—•—•—•—•—•—•—•—•—•—•

这条光谱线上，深黑与黑之间，乳白与白之间等仍然可以出现许多颜色，对色彩有特殊要求的职业人群（如染印行业、画家）对黑白之间的所有有区别意义的颜色都会给出一个特定的概念。显然，这些概念的数量可以是无穷无尽的，他们之间的关系是并列的，要像在这条光谱线上穷进所有事物就得把对每两个有区别意义的属性给出两个不同的词语概念。这种描述系统对事物属性无法作出穷尽的描写。

如果，我们把多元并列的属性描写变成下面这种二元的属性描写，情况就不一样了。

(3) 属性 X——[+ X]
　　　　　　[− X]

比如说，当 X 为“长”时，某物便可以说成具有［+长］或［−长］的属性，那么，所有的事物便可以用［+长］或［−长］来描写，进而可以用“长”这以概念穷尽所有事物：在所有事物中，某物要么“长”，要么“不长”，在“长”与“不长”之间没有任何东西，因为不存在“既长又不长的”的东西。同样，我们也可以赋于“黑”这一个概念两个不同的值，来穷尽所有的事物：某物要么“黑”，即［+ 黑］，要么“不黑”，即［− 黑］；具有“黑”属性的和具有“不黑”属性的事物加在一起肯定等于世界上所有的事物。用“深黑”来说，世界上的事物要么“深黑”，要么“不深黑”，“深黑”的加上“不深黑”的等于世界上所有的东西。注意，我们也可以使用“短”这个概念，那么世界上的事物要么“短”，即［+短］，要么“不短”，即［− 短］；和使用“长”时一样可以穷尽事物，但是不能同时使用“长”和“短”这两个概念，就是说，在同一个系统中不能同时出现［+长］、［− 长］、［+ 短］、［−短］。因为，［+ 长］等于［− 短］；［− 长］等于［+ 短］；除非“长”和“短”是两个并列的概念，就是说［+ 长］不等于［− 短］，［− 长］也不等于［+ 短］。不然的话就会回到了多元属性描写法上去了。让我们看几个语言研究中的例子。

对词汇和其他语言成分进行语义描写是所有语言理论都不可缺少的部分。这种语义描写多半采用对“语义特性”(Semantic Features) 进行刻画的方法。比如说，在早期的西方语言学理论中经常出现“有生命”(animate)、“无生命”(inanimate)、“阴性”(female)、“阳性”(male)、“成人”(adult) 等用来进行语义特征刻画的概念。使用二元属性描写的方法，我们可以建立起下面这种二元制的描写体系：

(3)	特征名称	有值特征
	[有生命]	[+ 有生命]
		[− 有生命]
	[阳性]	[+ 阳性]
		[− 阳性]
	[成人]	[+ 成人]
		[− 成人]

也可以用下面的二元制体系，它同 (3) 的功能相同：

(4)	特征名称	有值特征
	[无生命]	[+ 无生命]
		[− 无生命]
	[阴性]	[+ 阴性]

[－ 阴性]
[未成人]　[＋ 未成人]
[－ 未成人]

二元属性描写系统可依据（5）这个公式建立：

(5) [＋/－X]，其中 X 为任何一个表示属性的概念。

比如说，我们描写实词词项的语法范畴属性时，就可以出现以下几种可能：

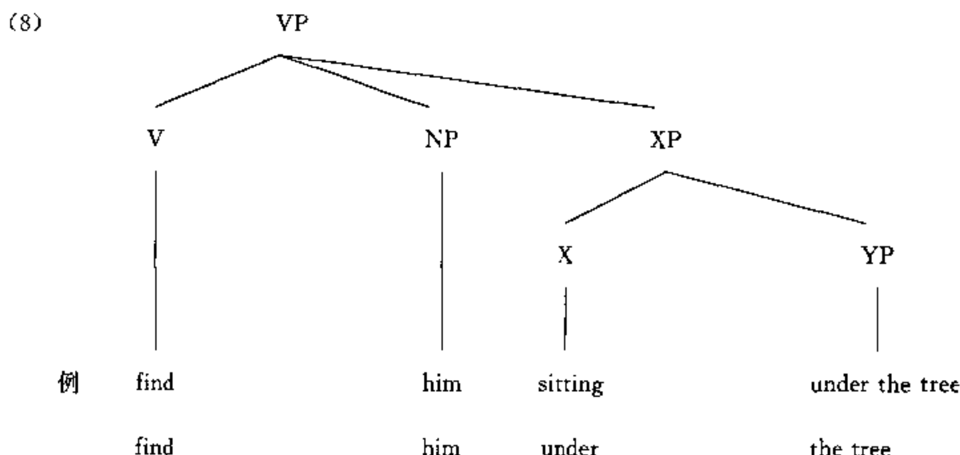
(6) a. [＋/－ 名词]
b. [＋/－ 动词]

按照这个二元制方法，一般所说的名词便具有 [＋名词] 的属性，其他包括动词、形容词、介词等词类在内的所有词项都具有 [－名词] 的属性；动词具有 [＋动词] 的属性，其他包括名词、形容词、介词在内的所有词项都具有 [－ 动词] 的属性；不管是以“名词”还是以“动词”的概念都可以穷尽所有的词类。但只靠“动词”“名词”这两个概念不足以描写出词项的句法特征。多个单一的有值概念可以组合在一起变成一个有值的概念群（cluster）。例如从（6）中的单一有值概念中可以得出下面四个有值概念群来：

(7) a. [＋ 名词，－动词]
b. [－名词，＋动词]
c. [＋ 名词，＋动词]
d. [－名词，－名词]

[＋名词，－动词] 和 [＋名词] 同效，可用来代表所有的名词；[－名词，＋动词] 和 [＋动词] 同效，可以代表所有的动词；[＋名词，＋动词] 可以代表所有的形容词，[－名词，－动词] 可以代表所有的介词。这种刻画词类句法属性的二元制方法实际上在西方现代理论语言学中得到广泛的应用。表面上看，这种二元的句法属性描写法于多元并列的语法属性描写法没有什么差别，说名词是 [＋名词，－动词] 与说名词是“名词”没有什么差别，说介词是 [－名词，－动词] 与说介词是“介词”没有什么差别。实际上，这两种描述法是有差别的。假如我们的理论需要把所有的实词从句法属性的角度无一遗漏地分成几类，就是说最后分成的类别能无一例外地把实词都囊括进去，有两种供我们选择的方法。一是用一套并列的概念（动词、名词、形容词、介词、副词等），把词分成相应的五类。那么，在句法操作规则中，就得分别说明在什么情况下可以使用动词，在什么情况下可以使用名词，在什么情况下可以使用形容词等等。如果遇见需要讲明在什么情况下只能使用名词，就得在规则陈述中加上“只能”二字加以限制，如果遇见既可以使用动词又可以使用介词的情况，就得在规则中讲明既可以使用动词又可以使用介词。这样，一方面规则不但显得繁杂，另一方面这种既可以这样又可以那样的特性在句法的结构描写式中无法得以表现。比如在（8）这个结构描写式

中就无法用一个概念表现在 X 这个结构位置上既可以出现动词又可以出现介词的特点：



最好的办法是给 X 这节点起一个名字，叫 [动词或介词]，这实际上正好是二元制中 [—名词，—动词] 这个句法属性群。如果，我们还保留句法结构中 N、V、P、A 这些标识符不变，而依据 (7) 中二元制的句法属性体系便可以简单说成 X 这个结构位置可出现具有 [—名词] 属性的词，即动词和介词，因为动词再 (7) 中被描写为 [—名词，+动词]，介词被描写为 [—名词，—动词]，两者共有 [—名词] 的属性。显然，这要比多元的方法来得简单明了。

在西方现代理论语言学中使用二元制属性描写法的实例有很多。而且，大多情况下是用在对数据性的词语进行特征刻画上，也有用在规定其他语言属性或属性群上的。比如，在语言类型学范围内，就有人用 [+/— 中心词在先] ([+/—head-initial]) 或 [+/— 中心词在后] ([+/— head-final]) 这一关于语言词序属性概念把语言分成两类：任何一种语言要么属 [+中心词在先]，要么属 [—中心词在先]，别无其他的归属。也有人用 [+/— 组列性] (+/— configurational]) 把人类语言分成 [+ 组列性] 和 [— 组列性]；一种语言要么属 [+ 组列性]，要么属 [— 组列性]，别无其他的归属。也有人用 [+/— pro 脱落] 这一二元属性，把语言分成两种：要么是 [+ pro 脱落] 的语言，要么是 [— pro 脱落] 的语言；也有人用 [+/— 疑问词移动] 这一二元制属性把语言分成 [+ 疑问词移动] 语言和 [—疑问词移动] 语言等等。这些简单的二分法，一分到底，在两者之间没有第三，第四种可能，实现了穷尽的目的。至于说用什么概念，那是理论方法之外的事。

运用二元制属性描写法的技术要点是，首先得把句法结构中的某一标识符当作一个句法属性概念 X ，然后赋予这个概念一个正值，一个负值，形成二元制的属性 [+ X] 和 [— X]，再用这两个有值的属性去刻画词项，结果就可以万无一失地把所有词是否能出现在 X 这个位置上描写出来。二元法在用来刻画有限的句法属性时，除了穷举至尽的优点外，在理论规则表述上还有着明显的简明性。二元法的另外一个优点表现在对语义特征的刻画描写上。句法特征和语义特征的一个差别在于，前者数量是有限的，概念基于形式化的结构之上，因而是明晰的，而后者究竟需要哪些概念、什么概念，十分含混不清。二元制的语义特性描写法虽然不能完全克服语义特性描写的根本困难，用什么概念描写语义特征合适虽然也不能完全求助于二元法，但它还是有其独特的理论意义。

4.7. 空范畴研究

在几乎所有的西方现代理论语言学的语法系统中都有关于空范畴问题的研究。空范畴(Empty Categories，简称 EC）是一个很复杂的技术用语。一般说来，空范畴是一个用来表示“没有声音形态、却有语义内容的结构成分”的形态标示符。空范畴在各家语法系统中的分类和使用意义不完全一样。研究得最详尽的要算生成语法。下面就以生成语法为例讨论一下什么是空泛畴以及为什么要研究空范畴的问题。

在生成语法的“管约论”和现行的“最简语言理论”系统模型中，有以下几种空范畴标示符，分别用来表示不同的结构成分：

（A）PRO

表示出现在不受支配但有语义角色标记的没有声音内容的结构成分。在这种定义下，最典型的 PRO 见诸于英语不定式谓语的主语位置：

（1）$John_i$ tried PRO_i to come early.

（2）$John_i$ promised Mary PRO_i to come early.

（3）$John_i$ persuaded $Mary_j$ PRO_j to come early.

（4）$John_i$ likes PRO_i reading books.

这种 PRO 必须带有一个表示所指意义的下标。而这个下标又必须和句子内某一先行成分同标共指。这种同标共指的关系叫“控制”(control)。在（1）中，PRO 表现出以下几个特征：(a）带有下标 i；(b）这个结构成分没有声音形式，当有声音形式时便成为一个不合语法的句子：

（5） * John tried Mary to go.

(c) PRO 所在的位置是一个有语义角色的位置，因为动词“to come”是一个“实义动词”，使用这个动词时必须有一个可充当其动作执行者这一角色的名词性成分，如果动词不是“实义动词”，也就是说动词不需要动作执行者这一语义角色，则 PRO 不能出现在这个位置上：

（6） * John tried PRO to seem to be happy.

(d) 与 PRO 相关的动词必须是不具“支配”(government）能力的动词不定式或动名词形式，PRO 在动词的主语位置上。在受时态成分“支配”的主语位置或受及物动词“支配”的宾语位置上则不能出现 PRO。

（7） * John persuaded Mary PRO would come.

(8) * John bought a book for Mary to read PRO.

(B) PRO_{arb}.

PRO_{arb}所标示的成分与 PRO 基本相同，但 PRO_{arb}不需有先行成分“控制”它，而独自带有一个下标 arb (arbitrary) 表示任指的“人们”之意：

(9) It is impossible PRO_{arb} to have him back.

(10) PRO_{arb} Smoking is harmful.

(C) t

t 的全称是 trace，标示在推导过程中移动成分移动后在原来位置上所留下来的“踪迹”，分两种：(A)“名词踪迹”(NP-trace) 和 (B)“疑问词踪迹”(Wh-trace)：

(11) $John_i$ was arrested t_i.

(12) $What_i$ does John think that Mary likes t_i?

(D) pro

pro 所表示的是受支配的、没有声音形式的、但又不是因移动而产生的结构成分，常见于动词曲折变化丰富的语言和“话题突出”(topic-prominent) 的语言中，如 (13) 这句意大利语和 (14B) 中的汉语：

(13) pro parle.

(14) A：张三来了吗？

B：pro 来了。

(E) 作为 vbl 的 t

vbl 的全称是 variable (变体)，标示“wh 词”或量词移动后在原来位置上留下来的踪迹，话题化产生的空范畴也属此类，这类空范畴也记作 t ：

(15) $what_i$ did John buy t_i.

(16) [[Every man]$_i$ [t_i [a woman]$_j$ loves t_j]].

(17) $That_i$ I don't know t_i.

(F) GP (寄生性空范畴)

GP 的全称是 Parasitic Gap，标示那种依存于其他类型空范畴的无声有义的结构成分：

(18) This is the book $which_i$ I returned t_i without reading GP_i.

(G) 空话题及空运算符 (Empty Topic, Empty Operator)

标示空话题的标示符是［TOP e］；标示空运算符的标识符是 0。［TOP e］常见于（19）这类汉语句子中，0 常见于（19）这类句子中：

（19）［TOP e］$_i$，我不认识 t_i。

（20）John$_i$ is too stubborn 0_i PRO$_j$ to talk to t_i.

（19）中 t_i为一变体性的 t，［TOP e］不是移动什么成分后产生的，因为在这句中不见有什么成分移动过。［TOP e］的存在有两个意义：一是我们讲“我不认识”这句话的语感包含有“有某一位特定东西或人我认识”的意思，二是为 t_i 提供一个先行成分去“支配”它。（20）中的 0 近似于（19）中［TOP e］的情况，它的存在只是为 t_i提供一个“支配”它的先行成分。以上七种空范畴是绝大多数语言学家公认的在语法系统中应该存在的空范畴。此外，有人认为还应该加上一种叫 Pro 的空范畴。这种空范畴见于下面的汉语中：

（21）a．［［Pro$_i$眼睛］大的］［姑娘］$_i$

b．［［Pro$_i$爸爸］有钱的］［孩子］$_i$

式中，“眼睛”与“姑娘”之间，“爸爸”与“孩子”之间都存在着“所属”与“被所属”的关系。

从以上含有空范畴的实例中可以看出，有的空范畴是直接描写句子意义的需要。比如，说“John tried to come early.”的意义结构描写式含有空范畴 PRO，是因为这句话本来就包含有动作“to come”的执行者和“tried”的执行者都应该是“John”这一个人。在“It is impossible to have him back.”一句中，我们的语感告诉我们“to have him back”的执行者是泛指的“人们”；在“This is the book which I returned without reading.”中，我们的语感也告诉我们“I returned”的东西是“the book”，“without reading”的东西也是“the book”，“来了”的执行者肯定是上文中的“张三”，等等。这些不言而喻的句子成分全都在我们的语感之中。为把这些没有声音形态但有实在意义的成分表达出来，语言学家们使用了空范畴。另外一种空范畴，在我们的语感中并没有直接觉察到。比如，我们说“John was arrested.”这句话时，不会像讲“John tried to come early”那样直接感觉到有什么空范畴的存在。那么为什么也要在这类句子结构描写式中加上一个空范畴标示符呢？原因同样也是意义描写上的需要。用一个空范畴符号 t_i，放在“arrested”的后面的位置上，由于这个位置毗邻这动词，t 提可以充当“arrested”的动作执行者的语义角色。又由于 t_i 和句子主语“John”同标共指，这种形式化的表述便描写出了“John”虽然处在远离动词“arrested”的主语位置，但我们仍然知道它和在“Someone arrested John.”里一样都是动词“arrested”的动作受事者这样一种语感。Every man loves a woman 这串声音文字符号承载着两种不同的意思：一种意思可以解释为“每个男人都有一个女人由他去爱”，另一种意思可以理解为“有那么一个女人，所有的男人都爱”。这两种不同的语义解释应分别表达成不同的描写式，关于前一种理解的形式化描写式是（22）的样子：

（22）［Every man］$_i$［t_i［a woman］$_j$ loves t_j］］

使用空范畴 t 是为了满足这种描写式的需要。直接感到有的空泛畴和结构描写上需要的空范畴实际上都是意义描写的需要。生成语法对这些空范畴类型做了很严格的区别，对每一种空范畴的“放行条件”（licensing condition）都做了细致而严格的规定。围绕着空范畴研究的论文和专著接连不断地出现，所涉及的语言也空前之多，一直是语言学家们热心探讨的重要课题。如此重视空范畴研究究竟有什么意义呢？

第一，西方现代理论语言学认为空范畴这种没有声音却有意义的语言成分对认识人脑的属性有特别的意义。如果能对空范畴做出正确的分类，并且对每一类空泛畴的存在给出精确的条件，就有可能得到通过单纯研究有声音形态的语言成分所得不到的有关人脑属性的认识，因为空泛畴是无声有义，人类的思维意识活动也是在不声不响地进行着。

第二，空范畴是在人类各种语言中普遍存在的现象，似乎是人类语言必不可少的东西。为什么人的语言一定有空范畴呢？为什么人脑“造出来”的语言一定离不开空范畴呢？为什么任何一种语言中相应于（23a）这句含有空范畴的汉语句子都必须含有空范畴，而任何一种语言都不允许出现对应于（23b）的句子呢？

（23）a. 张三逼我 PRO 去

b. * 张三逼我我去

也许，认识了空范畴的种类，找到了空范畴的句法分布及其存在的条件，也就对这些问题做出了回答。

第三，空范畴的句法分布情况在各种语言中几乎完全对应，语言 A 在 X 位置上的空范畴会对应于语言 B 在 X 位置上的一个空范畴，A 在 Y 上的空范畴对应于 B 在 Y 上的一个空范畴。如果事实确实如此的话，空范畴可能就是通向普遍语法的“直接通路”，从而解除了显性成分是否有普遍语法意义的困惑。

第四，在一种具体语言中，空范畴也是反映句间相关性的接口，这样，就有可能按照使用空范畴的统一条件把一种语言中的可能句子组织成一个有机的系统。

第五，推导过程所需要的空范畴有连接表达式与表达式的接口作用，从而获得系统的操作性。

第六，单纯出于语义解释需要而引入的空范畴是连接句法系统和其他形式语义系统（如一阶谓词逻辑）的接口。

第七，表示空话题的空范畴可能是连接“句子语法”（sentence grammar）和“篇章语法”（discourse grammar）的接口。

总之，研究空范畴具有研究其他成分不能得到的理论结果。

4.8. 西方现代语言学与其他语言理论的关系

西方现代理论语言学在四十多年的研究过程中已经形成了自身独具的解释性理论特征和方法论特征。虽然它是在批判传统的描写性语言理论基础上建立起来的，但是它并不排斥现

代的描写性语言理论，更不排斥实验语言理论。凡是以人脑与认知作为理论对象的，西方现代理论语言学家都看成是从事共同事业的同伍。这一点在下面关于理论语言学与其他语言学理论关系的简单说明中可以看得很清楚。

比如说，语言习得理论问题一直是心理语言学和实验心理语言学研究中的一个重要的理论主题。这个理论主题可以说首先是在理论语言学中形成的。1957 年 Chomsky 在他的《句法结构》一书中批评了行为主义的语言学习观，提出了关于“语言习得机制”(Language Acquisition Device, LAD) 和 “评价过程”(Evaluation Procedure) 的两个著名论断，而后引起了美国语言学界和心理学界的普遍注意，进而开创了心理语言学研究的一个新方向，使语言习得理论成为一个独立的语言学理论。理论语言学中提出的关于语言习得问题的一些重要理论假说便成了心理语言学中的研究课题。在心理语言学领域中人们开始通过各种实验心理学的方法观察研究儿童语言习得的心理发展过程，回答和验证人们在理论语言学中提出的假说。在这方面比较有影响的研究涉及到以下几个理论问题：(1) 有关语法独立性的问题；(2) 有关“代词化”(pronominalization) 的问题；(3) 有关“约束论”的问题；(4) 有关“定语从句化”(relativization) 的问题及 (5) 有关语言习得的“子集合论理论”(The Subset Theory) 的问题；(5) 有关语言习得中是否有“负面证据”(Negative Evidence) 的问题。

Chomsky 的“语法自治论”在引起一般理论语言学家注意的同时也引起了心理语言学家们的注意。他们通过实验观察的方法发现人脑的某一特定部位受到抑制后，受试者讲出的句子都是合乎语法的句子，但有许多合乎语法的句子与常识世界完全脱节，进而从实证的角度佐证了语言学研究从理论语言学的角度提出关于句子中的代词是从词库中的名词经过一个“代词化”的推导过程而来的观点。这个观点也引起了心理语言学家的兴趣。他们通过对幼儿初学语言期的跟踪调查，发现幼儿开始使用代词的时间在有的案例中早于名词的使用，在有的案例中晚于名词的使用。结果，无法证实理论语言学中的“代词化”过程究竟是语言习得的实际发展过程还是语法系统中的一个没有现实意义的推演过程。理论语言学中的“约束论”提出了关于反身代词和人称代词句法分布的条件，心理语言学家也是采用了实验观察的方法跟踪了英语儿童和汉语儿童学会使用反身代词(英语中的“himself, myself, yourself”和汉语中的“自己、他自己，你自己”) 和人称代词 (英语中的“him, me, you”和汉语中的“他”、“我”、“你”) 的顺序，结果发现与“约束论”相矛盾的结论。对此，理论语言学家们做出了两种不同的反应，一种人认为“约束论”和其他一些理论语言学中的理论一样，不是真实的心理过程和条件，因而具体的局部理论没有现实性的真伪可言；另一种人开始从另外一个角度解释“约束论”原有的条款并认为心理语言学上的所见与理论语言学原有的主张没有相悖之处。特殊疑问句的问题一直是理论语言学家们热衷于研究的问题，在心理语言学界也引起了很大的兴趣。不过，不同于理论语言学家，心理语言学家的兴趣不在于探讨什么形式化的表达更为理想，而是想从心理语言学的角度研究儿童语言发展过程中掌握特殊疑问句的时间顺序，从中找寻不同特殊疑问句间习得和存在上的相互关系。研究方法是观察实验和统计。结果发现，幼儿在学习使用特殊疑问句时，无一例外的都是先能说出对句子宾语发问的特殊疑问句，而后才会讲对句子主语发问的特殊疑问句，然后才能说对地点、时间、方式和原因发问的特殊疑问句。这个观察实验结果支持了理论语言学中关于“主要语义角色”和

“次要语义角色”的论断[①]。

语言习得研究中影响最大的理论要算一种称作“子集合论”的语言习得理论。这个理论的提出生动地说明了理论语言学研究和心理语言学研究相辅相成的关系。在语言习得问题上，理论语言学家们认为，一个儿童生下来的时候就有一个可对后天环境中的“未经分析的语料”(unanalyzed data) 进行加工处理的解释性系统，这个系统对所有的儿童都是一样的，加工处理的结果就使这个儿童所听到的语言（称作“外化语言”externalized language)“内化”了 (internalized)，这个内化的语言便是这个儿童母语的语法。用一个近似的比喻说，这个儿童好像一生下来的时候，就带着关于什么是可能语言的限制以及在这个限制内学会任何一种语言的可能，一旦他听到的语言和他所固有的一个可能相匹配的时候，他便掌握这种语言。对于这个理论思想，心理语言学家们立即想到这样一个问题：如果这个儿童所听到的语料同一个范围大一点的可能相匹配同时又和一个范围小一点的可能相匹配，那么这个孩子究竟是把那个大一点的可能内化为他的母语语法还是把那个小一点的可能内化为他的母语语法呢？带着这个问题，语言学从心理语言学的角度做了大量的观察实验，比较一致地得出了一个有名的语言习得理论 —“子集合论”，认为如果儿童所听到的语料中一部分材料为另一部分材料的子集，而这两部分都符合普遍语法的规定，那么前者应成为这个儿童母语语法的一部分。

随着研究的深入，普遍语法的内容也越来越丰富，心理语言学研究的课题也越来越多。近年来最引人注目的是 David S. Lebeaux 的博士论文《语言习得与语法形式》(Language Acquisition and the Form of Grammar)。在这篇论文中，Lebeaux 从心理语言学方面研究了 Chomsky《管约论》的语法系统结构，得出了下面几个心理语言学结论：(1) 词项确有树形结构；(2) 词项结构与句法结构有对应关系；(3) 存在着一个对应于儿童“电报化语法”(telegraphic speech) 的“纯语义关系”的表达层次 (representation of pure theta theory)；(4) 附加成分不是短语结构的组成部分，而是通过“移动”后加在短语结构上的；(5) 语言习得各阶段的划分与“深层结构”——“表层结构”之间的推导层次关系相吻合。这些通过观察实验得出的结论引起了心理语言学界和理论语言学界的极大兴趣。

同时，实验心理语言学、神经语言学也通过对失语症患者、动物比较研究等在研究着理论语言学家们和心理语言学家们共同感兴趣的问题，为实现揭示人脑的奥秘工作着。

---

① “主要语义角色”主要指名词性的表示“施事”、“受事”、“经历者”(experiencer)、“述题”(theme) 等语义角色，在特殊疑问句中使用“什么”、“谁”、“哪”等疑问词，“次要语义角色”指非名词性状语性的语义角色，在特殊疑问句中用“（在）哪里”、“什么时间”、“用什么”和“为什么”等表示。

中 篇

描写方法篇

5. 语言学中的人类学传统

5.1. 语言学在人类学中的地位

人类学有四个分支：人类生物学、考古学、人类文化学（民族学）和人类语言学。而在现代人类学研究中，语言学的地位越来越重要，因为语言是无所不在的，而且是人类独有的，是把人类和其他动物区别开来的主要特征。其他动物（包括灵长类动物）也进行交际，但只有人类使用语言① 来进行交际。生物学、解剖学和神经学的研究成果说明，语言的使用是人类长期进化的结果。人类的许多行为都是以语言行为为基础的，如：协作进行狩猎、耕种和运动，计算亲戚关系，安排婚姻，举行宗教仪式，组织军事远征等等。作为一种交际系统，语言比其他动物的交际系统更为精细、更为复杂，传递了其他交际系统所不能传递的信息。这些信息就是文化。和语言一样，文化也是人类的一种独特的现象。人类对他们所处的环境进行分类，并且对他们的日常生活赋予各种意义和动机，甚至创造了不存在的神祇和妖魔鬼怪，用不同的方式去认识自然的神秘力量。每一个人类社会都有其自身的文学、哲学、神学，没有语言，它们都不可能存在。为了描写人类的群体和它们的社会行为，人类学家都必须研究语言；因此语言研究就成为现代人类学的一个重要的分支。人类语言学和其他的专业一样，有它自身的研究方法、分析程序、术语和概念。

从进化论的角度看，人类学对语言的生物学基础很感兴趣，因为这是语言共同性② 的重要方面，但是不同的语言也有其不同的个性。世界的语言繁多，最低的估计是 2000 种，最高的估计是超过 5000 种。Crystal (1987) 列举了使用者超过一万人的语言，差不多有 1000 种。语言能力是物种遗传的，但作为一个客体的语言却是代代相传的，是一种学然后知之的行为；和文化系统一样，语言视时间、环境、需要的变化而有所变化。因此不管其对人类学的兴趣如何，大多数的语言学家都把他们的一部分精力花在研究、分析和描写某些语言的结构和内容。这种方法认为语言是独立于其他行为系统的系统，可以专门进行描写，因此又可称为描写主义或结构主义的方法。

5.2. 人类语言学的发展

社会语言学和文化语言学（ethnolinguistics）是在人类语言学的基础上发展起来的当代的新兴语言学学科，吸引了很多语言学家的注意力。它们继承了人类语言学研究方法的传统，结

① 指严格地按现代语言所定义的“语言”。

② linguistic universal 在本书中有几种译法：“共项”、“语言普遍现象”、“语言共同性”，国内也尚未统一，我们也不去统一，但希望读者留意。

合语言的各种社会因素和文化系统（如地区、民族、性别、年龄等等）来描写语言。

人类学家强调现场调查[1]，要作实际调查，就必须懂得所调查的民族和社区里所使用的语言。有些语言有现成的语法、词典和其他的资料（甚至录音材料），可为他们的研究提供帮助。只要掌握一些基本的词语和用法，就能通过接触来收集第一手资料。但是学另外一种语言终归要花时间，于是有的人只好求其次，寻找一些懂得他们所使用的语言和所调查的语言的操双语的资料提供人（informants），但是这究竟是隔靴搔痒。从人类学家的角度看，他们所感兴趣的并非语言的词语和语法本身，而是这些语言所包含的文化和社会色彩。可是语言又是载体，语言研究和文化研究的原始资料，几乎都是一样的。所以人类文化学家要想研究文化和社会，就离不开语言。

5.3. Malinowski 的贡献

对语言感兴趣的人类学家众多，在欧洲以 Bronislaw Malinowski（1889—1942）为代表。18 世纪欧洲经历过中世纪禁锢的封建生活后，开始对广大市民层关注（如莫里哀戏剧里聪明而狡诈的仆人，博马舍尔的《塞维勒的理发师》，英国 18 世纪的小说等等）。格林从德国农民的民间传说里获得创作灵感，更说明乡村的白丁也具有真正的美学价值。到了 19 世纪，注意力又移向工人阶级，甚至殖民地的非洲和亚洲里的底层。体现了这种注意力的转移的是人类学的兴起，人类学家对这些老百姓的详尽的描述导致了文化人类学的产生[2]。于是人类学家派遣他们的学生去作实地调查，而不依赖那些书面材料。1914 年，在牛津大学读人类学的 Malinowski 被派到英属 Trobiand 群岛，进行文化人类学研究。他是奥匈帝国籍的波兰人，被英国殖民地官员怀疑为间谍，不让他离境，他只好在群岛多呆了一倍的时间。Malinowski 在被扣留的时间里发展了一套观察原始社会的日常生活的调查方法，把长期的参与观察和敏感的访问结合起来，调查不但记录了 Trobiand 居民的明露的文化知识，而且还推断了他们的隐含的文化知识，具有深刻的洞察力，他在 1922 年所发表的报告被认为是文化人类学中的一场革命。

他在《西太平洋的亚尔古英雄》(Argonauts of the Western Pacific，1922) 里指出，他的研究采用了三方面的数据：

1. "氏族组织及其文化的解剖。" 这包括了像婚礼、家庭组织、亲属关系图、耕地和收成和测量等等。

2. 对"实际生活的难以测量的事物"的观察，如在现场调查日记里使用逐日记录的方法，对某些具体的行为作详细的描述。

3. 用本族人的术语、原话来记录"民族文化的陈述，有特征的记事，典型的话语，民间传说和咒语。" Malinowski 认为，作为本地人思维方式的文件（补充以自己的文化研究），这些资料可以成为别的兴趣不同的研究者进一步研究和分析的基础。这些语言数据的发表是十

---

[1] field work，亦可称为野外作业，但调查不一定是在"野外"。

[2] ethnography，一般译成"民族志"，ethnoi 在希腊语原有"其他人"，即非希腊人的"异族"的含义，Webster's New World Dictionary 把 ethnography 定义为 "the branch of anthropology that deals descriptively with speicific cultures，especially those of primitive peoples or groups." 以下凡有 ethn-开头的词语都翻成"文化"。

分有价值的，和收集其他语言研究与统计的数据具有同等作用。

Malinowski 的背景是德国的社会学说，这种学说在自然科学（Naturwissenschaft）和人文科学（Geisteswissenschaft）中划出一条界线：人文科学之所以异于自然科学是因为人类具有其他动物或无机体所没有的特点，它们能够产生和分享意义。因此要研究在一起生活的人群，就必须观察他们在社会中是怎样互相了解的。提出这种社会学说的是德国的历史学家和社会哲学家 Wilhelm Dilthey，他认为人文科学的方法应该是 hermeneutical，即解释性的，其目的是发现和传递所观察的人群的有意义的观点(meaning-perspectives)。Erickson (1990)认为早期的马克思也采取同样的立场，马克思虽然强调物质条件的决定性作用，但也关注那些足以决定有意义的观点的内容。后来的马克思主义者认为这些有意义的观点随着社会阶级地位的不同而有所变化。德国的社会学说和以 Comte 及 Durkheim 为代表的法国的社会学说不同，后者采取的是机械主义的、实证主义的观点，从牛顿物理学的角度去对待人与人的关系。

5.3.1. 语言环境

Malinowski 对语言学的影响至大，Firth (1957) 认为 Malinowski 对语言学的贡献是提供了“一种普遍的理论，特别是他使用了语境[①]和语言功能的概念。”这两个概念成为英国语言学家和语言哲学家的研究兴趣的中心。以 Firth 为代表的伦敦学派提出了语境论（contextualism)，而语境论的一个目标是对 Malinowski 的语言功能进行更精细的分类。按照 Malinowski 的看法 (1935)，基本的语义单位不是单词或单句，而是处在一些语言表达式的上下文中的句子，句子的意义体现为它在一个更大的意义整体中的功能。“在语言学中拓宽语境的含义很有好处，语境不但包括说出来的话，而且还包括脸部表情、姿势、身体活动，所有参与交谈的人和他们所处的那一部分的环境。”对语境的描写往往要求对使用该语言的社会有一个透彻的认识，例如在 Trobriand 语中有一句话 Bi-katumatay-da gala bi-giburuwa veyo-da, pela molu，这句话可译为(1)如果他们杀了我们，我们的亲人不会因为我们在饥荒死去而愤怒。或(2)他们不会杀我们，因为我们的亲人不会因为我们在饥荒时死去而愤怒。在这种语言里，gala 相当于否定式，它可以修饰前面的或后面的句子，故两种说法都是可能的。按照欧洲人的道德观，应为 (2)；但是从 Trobrian 人的道德规范看来，自己亲人在缺乏食物时被杀，应泰然处之，所以原话的正确理解应是 (1)。所以意义的分析离不开民族文化的描写。

5.3.2. 语言功能

Malinowski 的第二个理论支柱是语言的意义是在使用中体现的，因此语言是一种活动的方式，而不是反映的工具。他的观点可以表示为两个口号：“在行动中的语言”和“语义就是使用”(见 Leech，1977) 这些思想后来在 Wittgenstein 和 Austin 语言哲学著作得到进一步的发展，成为以 Halliday 为代表的系统功能学派的理论基础。Malinowski 认为有四种言语功能

---

① context 可以广义地理解为说话的环境，也可狭义地理解为一个句子所处的上下文。在汉语里，我们一般说“语境”，有时也说“上下文”，视具体场合而定。

以及与其语境相联系的四种意义：

1. 保持接触的意义(phatic meaning)，见于那些词语并不起到传递信息作用的言语环境，如“早上好”只是一种保持接触的手段，并非指那一天的早上很好。打个招呼完全是为了满足人群保持接触的需要，招呼里用什么词语是无关重要的。因此这种言语功能的意义和外部环境联系在一起。

2. 实际的意义（practical meaning)，这是言语在某一活动环境内的具体使用。词语的外部环境和实际的意义的关系是一种复杂的活动，而言语行为是其中的一个因素。这些活动包括人们对词语的反应，根据反应而作出的行为，这些活动最后产生的结果。在这些活动的背景下产生的词语都有实际的意义，才能达到这些活动所要求的目的。词语的实际意义是在劳动的共同协作中学到的。

3. 魔力的意义(magical meaning)，言语的使用者相信词语和外部世界具有一种神秘的联系，这不但见于儿童的语言习得，也见于原始部落和现代社会。例如在文明社会里的广告语言和政治语言。魔力的意义也是话语在特定的外部环境里的一种功能，这个外部环境对说话人神秘莫测，因此需要采取一些言语活动，以取得用正常途径所不能取得的结果。这些活动是以使用者相信其效力为前提的，所以在分析魔力的意义时，我们必须考虑使用者对魔力的词语所能产生的结果的信念。“对民族志研究者来说，魔力的词语有另外一种比它们的神秘效应更为重要的意义，就是它们对人类所产生的效应。”(1935) 由于相信了它们的魔力，人们就会采取相应的行动。所以信念和效应一起构成了使人类学家感到兴趣的魔力的意义。

4. 叙述的意义（narrative meaning)，只有在叙述的上下文的环境里，词语的意义才能表达思想。这种功能来自语用功能，因为人们必须在活动中，才能学会词语的意义，使用它们来说故事、描写事物、讲道理。这些在叙述的环境下所表达的思想，不仅是简单的陈述，还会通过词语的感情的力量，对听众产生效应。这就是日后 Austin 所说的表达性言语行为(locutionary act）和成事性行为（perlocutionary act)，不过 Malinowski 更为强调后者。

5.4. 法国的传统

在法国，把人类学和语言学联系起来的应归功于四位学者：语言学家 de Saussure 和 Meillet，社会学家 Durkheim 和人类学家 Mauss。de Saussure 是欧洲结构主义之父，在提出以语言系统为核心的共时语言学的独立地位、区分语言和言语等方面都对人类语言学的建立作出重大贡献。Meillet 是历史比较语言学家，他把语言学列为社会学的一部分，从社会阶层的划分的角度去解释语义的变化。在 Durkheim 社会学里，语言学是一门以社会事实为对象的独立的学问，它的核心问题不是社会中的活动者的有意义的观点，而是他们行为的“社会事实”。

5.5. 美洲的传统

在美洲大陆，对美国印第安语的兴趣也孕育着人类语言学的研究。研究各种印第安语的亲属关系有助于了解新世界的本地人的起源和特征，所以美国人类学家对研究美国印第安语

的分类怀有持续的兴趣。另外对语言范畴的起源和社会意义的解释也是美国人类语言学家长期关注的焦点，他们和法国学派的传统较一致，把语言和文化的统一看成是一种文化产物和社会传统，而不是一种事件或社会活动。美国人类语言学的另一个特点是注意文本的收集。

5.5.1. Boas 的贡献

现代美国人类语言学的奠基人是 Franz Boas，Boas (1858--1942) 原籍德国，年青时到加拿大研究爱斯基摩人，后又研究加拿大西部和美国的印第安人。他学过物理学和地理学，对生物人类学和统计方法的发展都作出过贡献。同时他又对语言、民间传说、艺术作过深刻的研究。这就使他能够把科学的追求和他的人文科学学识结合起来。他既是一个人类语言学的教育家，又是一个有实际经验的语言现场调查工作者。他在 Clark，Columbia 和 Brinton 大学训练了一批又一批的人类语言学家，包括著名语言学家 Sapir 和 Kroeber。Bloomfield 也承认 Boas 对他的影响，“在我们的工作中，我们把 Boas 看成是美国语言研究的先驱和大师，他是我们大家在这个或那个意义上的导师。”Boas 在大学里根据自己的经验设计了一套完整的、把语言学包括在内的人类学教学计划，并且发展了一套人类语言学的研究方法，教导他的学生怎样通过文本来分析语言结构。作为一个现场调查工作者，他的贡献在于首先强调每一种语言都应按照它本身的配置，而不能按照一个事先决定的框框（如对拉丁语法作一点修补）来进行描写。另外他又注意语法和词汇的研究必须辅以大量原文的文本。Boas 主编了《美国印第安语手册》，并撰写了一篇长达 80 页的《序言》(1911)。在《序言》里，他讨论了很多语言学的概念和方法，日后都得到了充分的发展。他认为语言研究对文化研究的重要性体现在两个方面：

1. 它是实际的需要。掌握一种本地人所说的语言就不用依赖翻译，或使用洋泾浜英语和其它的方法来和资料提供人进行交谈。Boas 主张，而且他自己也身体力行地对所有题目进行文本记录，都用本地语言来收集第一手资料。记录下来的东西日后可以进行翻译和分析。有些题目只能通过语言来进行研究，如诗歌、祈祷文、演说、人名和土名。他自己就独立地对 Kwakiutl 语和其他印第安语译成文字来进行分析；在他所组织的人类学课程里，也专门训练他的学生使用这种技能。他在《序言》里指出，“掌握语言是准确而透彻地了解［所研究的文化的］知识的不可缺少的手段，因为通过聆听本地人的交谈和参与他们的日常生活可以获得很多信息，是不懂得该语言的观察者所无法做到的。”

2. Boas 还进一步指出，从理论上看，语言和思维具有密切的关系。只有通过语言才能了解一些“无意识的现象”——如意念的分类，以及怎样用相同的或相关的词语，或用比喻建立联想的方式来表达这些意念。所以，无论是从实际或理论的观点来看，语言研究对人类学都是至关重要的，“一方面，没有实际的语言的知识不可能透彻地洞察民族学；另一方面，人类语言所表达的基本概念也和民族文化现象不可分割。因为从根本上看，语言的特殊特征是鲜明地反映在世界上人们的观点和习惯上面的。”

5.5.2. 描写语言学和结构语言学的诞生

Boas 的这些观点成为美国描写语言学和结构语言学的催生剂。他的学生 Sapir 是一个高

超的现场调查工作者和语言形式的分析家。他善于发现型式，进行细致的描写和比较，揭示了新大陆语言的亲属关系。Bloomfield 也深受 Boas 的影响，1925 年他在发起美国语言学会和《语言》杂志的文章中指出，"人类言语的直接观察和记录的工作和民族学的工作很相似；在我国，这方面的工作做得最好的是民族学-语言学学派。"Bloomfield 和 Sapir 一道共同建立了人类语言学中进行现场调查和描写语言的标准，奠定了使用印欧语的比较方法来研究美洲的无文字的语言的历史基础。Bloomfield 在他的早期著作里，还谈到认知范畴，但是在他 1933 年所发表《语言论》里，他把语言研究局限在行为主义的框架里，他提出与"心灵主义"相对立的一种"科学的"方法，集中在推行描写主义。他强调语言是控制别人活动的非言语行为是一种替代物，对 Malinowski 从社会人类学的广角所提到的各种语言功能及其文化联系，却绝口不提。所以 Bloomfield 对"语言"和"语言学"的理解是狭义的，与狭义的理解无关的东西都搁置在一旁了。但是在近年来，对语言功能的研究却脱颖而出，与布拉格学派有密切联系的 Roman Jakobson 对"语言"和"语言学"持广义的理解，而且他把自己的研究和美国的人类学联系在一起。他早期曾来美国和 Boas 一起生活了几个月，并且写过两篇介绍 Boas 的文章。

5.6. 人类语言学的特征

Hymes（1964）在回顾了英、法、美的几种人类语言学思潮后指出两个明显的特征：

1. 语言学和人类学有密切联系，这对两门学科的成长都有作用，对认识语言和语言描写的性质，对语言在文化中的地位都有影响。

2. 两个学科在当代的历史中是不可分的。将来的发展如何尚难以预料，但是两者在美国的联系不但会持续下去，而且还会有所发展，年青一代的人类学家接受了语言学方法的新发展，继续在世界的许多地区对描写语言学和历史语言学作出人类学的贡献，从社会学和心理学两方面来研究语言使用受到了空前的注意。年轻人类学家研究语言的主要兴趣是语义描述和社会语言学。这种兴趣中心的转移使 20 世纪的语言学历史分为两半：在头半个世纪，主要是追求语言学作为研究对象的独立地位和集中在描写语言结构；在后半个世纪，主要是研究语言在社会文化环境中的一体化和集中在功能的分析①。

从方法论的角度，我们不妨增加一条，

3. 人类学的研究方法促进了描写语言学和结构语言学的发展，而它们回过头来又充实了人类学的研究方法。这种方法的基本特征是采用定性的研究方法，但在定性研究的基础上又采用了一些定量的手段。

① 我们不要忘记 Hymes 是社会语言学家，他的归纳只是说明从人类学到社会语言学的发展，并不能概括美国语言学发展的全貌。

5.7. 人类语言学和定性研究方法

人类语言学使用的基本上是人类学和社会学的定性研究方法。这种方法也是描写语言学和结构语言学，乃至社会语言学（包括方言调查）、语言共同性调查、语篇分析分析方法，因此有必要结合语言学研究，对这种方法进行分析和讨论。人类学和社会学的定性研究方法研究在某一具体环境里的人类行为，这种行为必须是自然而然地在该环境下产生的，不受研究者在其中的作用所影响。这些方法力图从所观察的对象或群体的角度去提供数据，这样研究者的文化或智能方面的偏见就不会歪曲数据的收集、解释和提供。Jacob（1987）指出，所谓定性的方法其实包括了各种不同的研究方法，她专门回顾了 5 种：

5.7.1. 生态心理学

生态心理学（Ecological Psychology）的研究对象为自然产生的人类行为以及它和环境的关系。生态心理学采取了两种方法以描写行为和发现行为的规律性。一种是样本记录（specimen records），由熟练的观察家对处于自然的、未经事先安排的环境下的人（通常是儿童）的行为作一段时期的描写，先是记录行为的流程，然后把它分成单位，再进行分析。另一种是行为背景调查（behavior setting survey），研究的是与特定的时间、地点、事物相联系的稳定而客观的人际行为型式。

5.7.2. 整体文化人类学

整体文化人类学（Holistic Ethnography）通过描写群体的信念和活动来描写和分析一种文化的各个组成部分，以及这些部分怎样共同构成文化的统一体。整体文化人类学的基本方法是（1）研究者必须亲自参加现场调查以收集第一手的证据；（2）文化人类学家必须反映当地人的观点——“他对他的世界的看法”；（3）要了解这些看法必须要有原话实录；（4）采取各种方法去收集各种数据。Malinowski 和 Boas 就是这个学派的代表人物。

5.7.3. 文化语言交际学

文化语言交际学（Ethnography of Communication）研究一个文化群体的成员或不同文化群体的成员之间怎样进行社会交往。它感兴趣的是面对面的交往以及这些“微观的”过程是怎样和文化、社会组织的“宏观的”问题联系在一起的，所采取的方法是描写这些面对面交往的型式。例如采用录音或录像手段来记录具体的过程，然后进行数据归纳，提出一个编码范畴系统，以处理收集的语料。

5.7.4. 认知人类学

认知人类学（Cognitive Anthropology）也称为新文化人类学（new ethnography）或文化科学（ethnoscience）。认知人类学从认知心理的角度去研究文化：通过研究语义系统（最近还扩展到语篇）来了解文化知识的认知结构。它和语言学的关系至为密切，例如它采用了采样的方法去收集原话实录，然后进行各种形式分析，如领域分析（domain analysis），找出文化领域及其使用的词语；分类分析（taxonomy analysis），找出文化领域组织的方式；成分分析（componential analysis），找出每个领域中的词语的属性；主题分析（theme analysis），找出领域之间的关系以及它们怎样联系在一起，结成文化的统一体。

5.7.5. 符号连接主义

符号连接主义（Symbolic Interactionism）主张个人的经验是通过他自己对经验的解释来传递的，而这些解释产生自个人和别人的相互接触，并被用于达到某些特定的目标。符号连接主义认为人类异于禽兽，不是靠本能和条件反射来行动的，而是根据其所认识的事物的意义而行动的。但是意义是符号性现象，人类不仅生活在物质环境里，而且还生活在符号环境里，因此人类的行动不但是对物质的刺激的反应，而且也是对符号的反应。符号连接主义主张描写符号相互接触的过程，以连接了解人类的行为，因此它从传记、自传、个案、信件、访问等等方面去收集材料，以期深入到活动者的经验中去。

这 5 种定性研究方法其实都是或多或少地和人类语言学有关的，换句话说，它们离不开语言研究，都要通过语言来了解人的经验、观点、认知结构、相互接触、文化背景。其实它也是现象学（phenomenology）、解释学（hermeneutics）、系统论（systems theory）、混沌论（chaos theory）、定向定性探究（orientational qualitative inquiry）经常采用的一种研究方法（见 Patton，1990），限于篇幅，这里不再展开。

6. 定性研究方法

6.1. 定性研究方法的原则

6.1.1. 自然观察

定性研究方法的首要原则是研究者不操纵研究背景，而是把研究背景看成是自然发生的事件、过程、相互关系。它的目的是了解在自然发生的状态下的自然发生的现象。自然观察方法也就是以发现为本的方法：研究者事先不带有任何框框，尽量避免操纵所研究的背景，也不对研究结果提出任何制约。所以这种方法十分强调不干预性。我们所观察的语言和文化处于一个动态发展的过程，必须忠实地记录这些发展，收集各方面的数据。自然观察的方法也叫做现场调查研究（field research），现场调查研究收集的是定性的数据，因为有些观察难以归结为数字，例如一位研究者可能注意到某一部族首领在政治集会上有一种“家长式作风”，但这种作风不容易用定量的方式来表示。

和自然观察法很不相同的是实验方法，实验方法采用控制和操纵的手段，根据假设专门设计实验，使某些要观察的行为在实验室的环境下更为集中地显示出来。尽管自然观察和实验方法的出发点和所采取的途径不一样，但是它们可以互为补充，相辅相成。自然观察可用于考察现象，发现问题，从一般到特殊；而实验方法可以对所发现的问题进行集中和系统的观察，从特殊到一般。把两者的结果加以对照和比较，就能由表及里，深入事物的本质。在讨论定性研究方法时，我们也顺便接触到实验方法，以兹比较。在下篇里，我们还要专门讨论实验方法。

视观察者在现场调查中的地位和作用，可以分为直接的（非参与性的）观察、参与性观察和个案研究三种：

6.1.1.1. 直接观察

调查人完全不参与所调查的事件的任何活动，只是客观地观察整个活动的过程，例如在公共汽车里听乘务员和乘客的对话。萧伯纳的《卖花女》(Pygmalion）一开始就表现了一个语言学家采取这种方法来记录伦敦土话，受到误会。这种从旁观察的方法能够保证数据的客观性和准确性。但是所能获得的数据都是表面的语言材料，说话人的经验和感受则难以了解，所以观察人虽然不会影响所调查的事物，但却不易深入了解所调查的事物。

6.1.1.2. 参与性观察

观察者参与他所观察的活动，但有程度上的不同。一种是完全参与（complete participant）的观察，被调查的人不知道观察者的身份和调查目的，调查者参与所有他感兴趣的领域，和被调查人自然地进行交往。调查者可以是真心实意地、也可以装成真心实意地参

与活动，总之人们不知道他是一个调查者，而是一个参与者。这种观察方法的好处是能够以参与者的身份获得亲身感受，但参与活动本身往往会影响你所要观察的社会过程。另一种是作为观察者参与（participant-as-observer）的观察，参与所调查的全部活动，但让人们知道你是在进行调查研究。这种调查方法的问题在于被调查人可能把注意力转移到你的研究项目，而忽略了社会活动的自然性，使你所调查的活动缺乏典型性。还有一种是作为参与者的观察者（observer-as-participant）的观察，调查者的身份是明白的，他也参与被调查人的社会过程，而且并不假装他是真正的参与者。

6.1.1.3. **个案研究**

深入地调查一个案例，以充分地了解与这个案例有关联的各方面的因素，这就是我们通常说的"解剖麻雀"的方法。采取个案研究方法的出发点是多种多样的，有时候是为了研究某些特定的类型，如外语学得好的人采用什么学习策略？有时候是因为所研究的案例不多见，如大脑语言中枢的某些位置受到损害而影响了语言能力的病人。有时候是因为纵向观察的需要而集中在个别案例研究，如儿童语言习得的全过程。

6.1.2. 归纳分析

自然观察意味着对所调查的对象不能带有任何事先想好的框框，因此它是一种探索性的研究，必须采用归纳法的逻辑手段。实验方法采用了演绎的逻辑手段，可以说是检验假设的研究（hypothesis-testing research）；定性研究则不同，它从开放性的观察开始，事先没有什么模型或假设，也不规定观察什么变量，而是根据所收集的数据进行分析过滤，概括出一般的型式，它所包括的变量，以及这些变量的相互关系，然后才产生假设。可以说是产生假设的研究（hypothesis-generating research），归纳法和演绎法的不同可见于图 6.1：

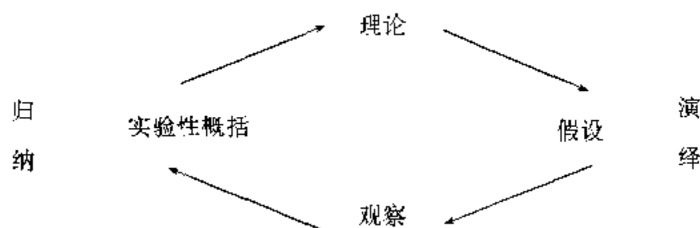


图6.1 演绎法和归纳法的差异

在科学研究中，往往需要交替使用归纳法和演绎法。在演绎逻辑阶段，我们推断出观察；在归纳逻辑阶段，我们又根据观察来进行推断。

但是它们的作用又有所不同，图 6.2 是比较两种方法的一个示意图，表示的是学习时间和考试中的级别的关系，学习时间多，级别也高。演绎法的过程是：假设—观察—接受或拒绝假设；归纳法的过程是：观察—寻找型式—初步结论。两种方法都离不开观察。

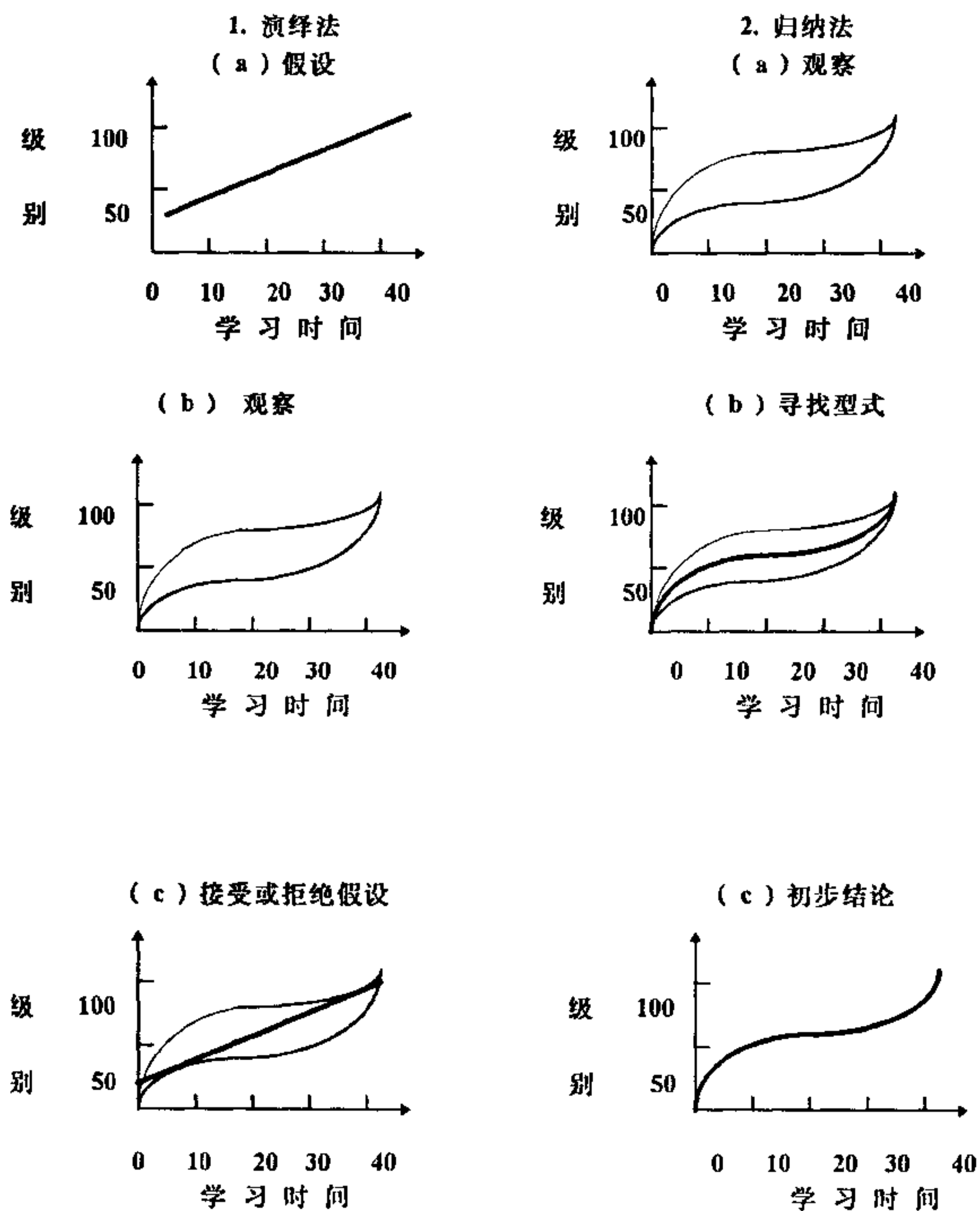

1. 演绎法
(a)假设
级 100
别 50
0 10 20 30 40
学 习 时 间
2. 归纳法
(a)观察
级 100
别 50
0 10 20 30 40
学 习 时 间
(b) 观察
级 100
别 50
0 10 20 30 40
学 习 时 间
(b)寻找型式
级 100
别 50
0 10 20 30 40
学 习 时 间
(c)接受或拒绝假设
级 100
别 50
0 10 20 30 40
学 习 时 间
(c)初步结论
级 100
别 50
0 10 20 30 40
学 习 时 间

图 6.2

6.1.3. 全面观

如上所述,归纳法是从一般到特殊的方法,因此它十分强调全面的观点。全面的观点假定所调查的事物是一个整体,是一个部分之和大于全体的复杂系统;它也假定必须把所调查的事物放在它的复杂的环境下来了解。定性方法所强调的全面的观点和定量的实验方法所使用的逻辑和程序很不相同。定量的实验方法需要对能够作统计分析的自变量和依变量[①]进行操纵,所得到的结果都是确定和测量出来的特定的变量。所调查的项目的参加者都是标准化的、量化的单位。所描述的变量有的是和调查的事物直接有关的,例如学生在语言水平考试中的分数和他的语言能力的关系;有的则是作为一个更大的模型的标志,和所调查的事物间接有关,例如通过了解犯罪率来了解社会的富庶程度。这些变量可以加以汇总,用线性关系来检验我们所提出的假设,并且推断出变量之间的关系。这种方法的基本点是(1)所调查的结果和处理方法可以分别用自变量来表示;(2)这些变量可以量化;(3)这些变量的关系可以用统计方式来表示。

从自然观察的定性方法论者的角度来看,实验方法的问题在于:(1)它把现实世界的经验的复杂性过于简单化了;(2)它忽略了那些不能量化的因素;(3)它不能作为一个整体去表示所调查的事物。所以单靠收集和测量孤立的变量的数据是远远不够的。全面的观点主张多方面收集数据,以求对所研究问题获得一个综合的、完整的了解。这意味着在收集数据的时候,所研究的每一个个案、事件、背景都应看成是一个独立的整体,都有其自身的意义和一系列的相关因素。Patton (1990) 曾经举了一个说明全面观的例子,一个葡萄牙人在他的国家边远地区开车旅行,在路上碰上一个牧羊人赶着一个相当大的羊群,挡住他的去路。他就下车和牧羊人交谈。

"你有多少只羊?"他问道。

"我不知道。"年青的牧羊人答道。

旅游者感到有点奇怪,就问他,"如果你不知道有多少只羊,你怎样赶羊?你怎么会知道少了一只羊?"

牧羊人被这个问题难住了,他解释道,"我不必算有多少头。我熟悉每一头羊。我熟悉整个羊群,如果它不够数,我马上就会知道。"

使用定量方法分离出来变量和标志,好处是经济可行、精密准确、易于分析,其信度、效度、可信度都可以量化,结论有力,令人信服。

使用定性方法来全面反映所调查的事物的背景和各方面的影响的好处是把注意力集中在事物的复杂性、因素的相互作用、环境的影响、独特性和精细差异。

人类语言学强调定性方法,因为它是下面几个概念的支撑点:

1. 语言和其他的行为是互相依存的,不能孤立地研究语言。
2. 使用语言的环境十分重要,这些环境大都和非言语行为有关。
3. 各种语言有很大差异性,只能对它们作具体的描述和分析,不能套用别的语言的模式。
4. 语言理论有可能从实际的现场调查和对语言功能的分析中产生。

应该指出,定性方法和定量方法也不是对立的,而是互相补充的。使用统计学的方法可

① 自变量(independent variables)也称为独立变量;依变量(dependent variables)也称为因变量。在10.4里会专门讨论。

以弥补定量方法的不足，故Boas在大学培养人类语言学家时，亲自开设统计学课程。很多文化语言学、社会语言学、语料语言学的研究都采用了统计方法。

6.1.4. 综合法

全面的观点隐含着综合法（synthetic approach）。我们所研究的事物都包括许多个组成部分，可采取不同的方法去对待它们。一种是综合法，它强调这些部分的相互依存关系，把这些部分看成是一个互有联系的整体。另一种是分析法（analytic approach），它强调这些组成部分在构成整体中的作用，着重观察在某一个平面上的一个或一组因素在构成整个系统中的作用。综合法也称为系统的方法。

试以第二语言习得为例，第二语言习得包括许多因素，每一个因素都可以是一个研究的领域。例如我们可以研究母语对第二语言习得[①]的影响，可以研究不同的学习者性格变量的作用，可以研究社会环境的作用和个人与社会环境的互相作用，可以研究语言学习的生理和生物基础，等等。这个清单还可以不断地延长下去，无穷无尽。系统的方法可以帮助我们把复杂性简约化。我们可以用某些统一的范畴把所有的因素组合在一起，例如"生物因素"、"语言因素"、"情感因素"，以便讨论其复杂性。每一个范畴都是第二语言习得的一个子系统，子系统中还可以包括许多子系统，如语言系统中还可分为句法系统、语音系统等等。这些子系统并非存在于真空之中，而是互相联系的，例如在习得元音的时候，其他方面的习得（如音节结构、社会语言学的语音变异）也同时进行，因而可能对它产生影响。所以第二语言习得研究是一些互相联系的系统，如图6.3：

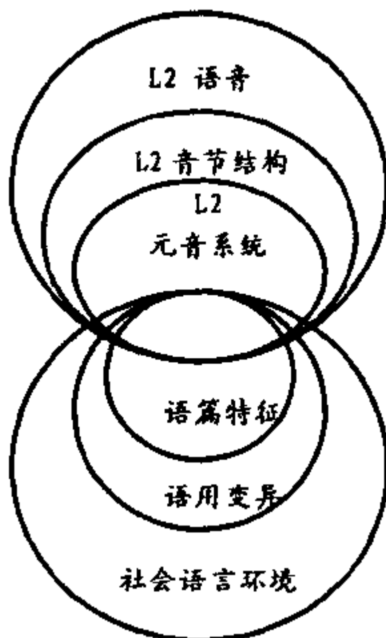


图6.3 第二语言习得作为一个互相联系的系统

---

① 习得（acquisition）原来是指在自然环境下学会母语，这个概念后来伸延到在自然环境（如语言社区）下学习第二种语言，更进一步伸延到非自然环境（如课堂里）学习第二种语言。习得强调的是学习者本身的学习。

现代各种流派的语言学家都赞成把语言看成是一个系统，结构主义的奠基人 Saussure (1916) 认为“语言是一个系统，它的任何部分都可以而且应该从它们共时的连带关系方面去加以考虑。”“语言是一个纯粹的价值系统，除它的各项要素的暂时状态以外并不决定于任何东西。”“语言既是一个系统，它的各项要素都有连带关系，而且其中每项要素的价值都只是因为有其他各项要素同时存在的结果。”Sapir (1921) 在他的《语言论》指出，语言是“一个约定俗成的任意性的符号系统。”Bloomfield (1933) 也认为，“语言作为一个可行的信号系统，只有数量不多的信号单位，而信号所表示的却是实际世界的全部内容，其差异是无穷无尽的。”Firth 则主张把结构和系统区别对待，结构是组合的、平面的、而系统是聚合的、纵向的。“系统为结构的因素提供价值，而系统的安排又依赖于结构。”(F. Palmer 所编的 Selected Papers of J. R. Firth, 1968, 138f) Halliday (1985) 还进一步提出系统功能语法，“语言中的每一个因素都可以根据它在整个语言系统中的功能来解释。”就连在方法论上不赞成描写主义者和结构主义者所使用的归纳法的 Chomsky (1965)，也承认生成语法“是一个规则系统，它用某种明示的、定义得很好的方法对句子进行描述。”

但是语言的系统观和语言研究的综合的、系统的方法虽有密切关系，却不是等同的；也就是说主张语言系统论的人不一定采用系统的（综合的、归纳的）方法。一般来说，人类语言学家、描写语言学家、结构语言学家较主张采用综合的、系统的研究方法，从不同因素、不同部分、不同层面去研究它们之间的关系；而理论语言学家（如 Chomsky）则采取演绎的方法来推导语言理论，Hjelmslev (1969) 也强调语言的系统观，但当他谈到发展语言理论时，却鲜明地指出，语言理论必须是演绎的，它是不断的分析，而不是不断的综合，“它根据它的前提推导出具有最大的概括性的各种可能性，然后应用到大量的实际数据中去。”“合适的程序应该是采取分析的和指定的，而不是综合的和概括的运动，从类到成分。①”

6.1.5. 分类系统

要对语言系统和语言结构的不同部分、不同因素、不同层面进行描写、比较和分析，从而找出共同性和规律性，甚至建立型式，就必须有一个严格的分类系统。这是整理和归纳数据的一个重要环节。使用分类系统方法的不限于图书馆学和生物学，它是人类思维的一个基本原则。分类系统可以根据数据来归纳，也可以由研究者根据别的信息源来决定，但是不管用哪一种方法，都应该能够经受别的研究者检验。历史比较语言学从纵向研究语言的谱系分类，类型语言学则从横向研究人类语言体系，其研究成果都体现为分类系统。

1. 类系统应该有几条共同的准则：

(1) 完整性。它必须能够解释所有的数据，能够准确而合适地反映所有的语言事实。也有人把这叫做详尽性 (exhaustiveness)。

(2) 一致性。它必须前后一致，没有内部的矛盾。一种数据只能放在分类系统中一个位置上。一些未经分析的原始数据必须能够完全检索出来。

---

① Hjelmslev 认为语言理论的目标不应是用归纳法来归类 (class)，而是用演绎法来发现最大程度上的语言的局部变体，所以说从类到成分 (components)。

（3）经济性。力图简洁，避免烦琐。各种标记必须清楚易记。

2. 分类系统还必须考虑三方面的关系：

（1）分子和同分结构的关系（allo/eme relation）。例如 allomorph 为词素变体，同一词素有不同形式，英语的复数词素为/s，z，iz/；allophone 为音素变体，同一音素也可以有不同形式，如/p/ 在 put 中是送气的，在 span 中是不送气的。

（2）类和成员的关系（class/member relation）。例如 fruit 是上义词，apple 是其下义词，而 Macintosh Apple 又是 apple 的下义词。

（3）结构和成分的关系（construction/constituent relation）。例如“小李看书。”是一个句子结构。这个结构可分为两个主要成分“小李”和“看书”；“看书”又可分为两个成分：“看”和“书”。

在建立分类系统时，还必须注意分析系统的外部结构和内部结构，这就是 Pike，K.（1967）所说的“etic 分类法”和“emic 分类法”[①]。etic 来自 phonetic（语音），而 emic 则来自 phonemic（音素），在语音分析中，语言学家尽量记录新语言的语音形式，不管它们在语言中有些什么联系。音素分析则牵涉到用语音数据去发现新语言的音素系统。所以 etic（外部的）分类法意味着对行为进行客观的外部的描述，不必作什么解释；而只有经过长期的观察后，人们才能由表及里，解释行为的 emic（内部的）含义。

Etic 和 emic 方法在研究语言和文化中的不同点有：

1. Etic 法从系统的外部研究行为（包括语言），它所使用的单位都是事先存在的（如代表各种声音的语音符号）；emic 法是从系统内部来研究，只有通过分析某一系统后才能发现 emic 的单位（一个语音学家在分析言语后才能了解音素）。Pike 认为，“Etic 系统是根据系统以外的范畴或‘逻辑计划’建立起来的，它也可以来自‘部分的信息’。”而“emic 范畴必须和系统的内部功能有关，要求通晓整个系统。”所以，“etic 数据是暂时的、初步的”，而“emic 数据则是精练的、最终的。”

2. Etic 法是跨文化的，可以比较的，它可以同时应用在几种语言和文化。而 emic 法则是针对某一特定的语言或文化，它只能在一时应用在一种语言（或文化）上面。“Etic 法的价值在于它能对世界的各种行为作出一个综观，更快地掌握一种新语言的数据；而 emic 法的价值则在于它表明一个语言或一个文化是一个有机的整体，使人们了解在生活舞台上的每一个演员——他们的态度、动机、兴趣、反应、冲突和性格。”

3. Etic 的范畴可以直接测量，因而是绝对的；而 emic 的范畴则是间接的，相对的。例如按照 etic 范畴，英语和捷克语都有［m］，［n］，［N］这三个音，但是从 emic 范畴看来，它们则是相对于特定的语言系统的。在英语里，这三个音素是不一样的，如 sum，sun，sung；但在捷克语，/m/和/n/是有区别的，而/ N /则是/n/的音位变体。另一种说法是，讲英语的人用 etic 法来进行信息编码（他们对某些 p 的送气强些，某些弱些，某些不送气，是无意识的），而在解码时则使用 emic 法，只注意那些显著性特征，例如靠浊音来区别 bunch 和 punch 中的/p/和/b/。

---

① etic 和 emic 很难翻译成对应的汉语，故保留其原文。如一定要翻，可称为“外部的”和“内部的”，但原来的一些意义便损失了。

Pike的这种区分法在人类学中得到广泛应用，例如考古学家Deetz，J.（1967）就建议在事实（facts）、事实素（factemes）和事实变体（allofacts）之间作出区别，为了了解这三个概念的区别，我们不妨看看他所举的箭头的例子：

箭头的凹口都是用来扣在箭杆上，因此它们都有同样的功能意义，可称为事实素（类似音素）。但是这些凹口都不相同：有方的，有三角的，有圆的等等，这些凹口的差异，可称为事实变体（类似语音变体）。而每一个凹口，尽管形状不一，都可称为一件事实（类似phone，单音）。



图6.4 事实素和事实变体

同样的，民俗学者Dundes（1962）也在主题（motif）的基础上，创造了主题素（motifeme）和主题变体（allomotif）这两个词语。例如世界上的民间传说都有一个主题素：英雄经历磨难或危险的考验。但是考验的方式很不一样，可说是主题变体。而每次考验的叙述都可以看成是故事中主题的出现。

但是，尽管这两种方法的区别在语言研究中是显而易见的，应用到非语言行为的研究上，却仍有不少争议，主要是围绕etic法的地位和作用而展开的。一种看法是etic法的描写无非是发现emic系统的一个前提，是一块踏脚石；另一种看法是两者都同样重要，不存在把一种描写转换到另一种描写的问题，我们可以从etic描写的角度去解释emic描写，也可以从emic描写的角度去解释etic描写。一份好的人类学报告应该既包括观察者尽量客观的描写（外部的etic描写），也包括本地的资料提供人在讨论自己的文化时头脑中的思考的描写（内部的emic描写）。

6.1.6. 动态的发展

这种观点认为我们所观察的事物是动态的、发展的，随着它所处的环境的变化而变化。因此定性方法描述和分析的不但是各种关系（静态的排列），而且更重要的是过程（动态的变化）的描述。定量方法也收集收据，以说明所测量的变量发生了什么变化，从而决定所调查的事物在产生这些变化中的价值和作用，所以定量方法较适合于研究那些较稳定的、较一致的、能够辨认出来的处理方法和能够量化的结果，以比较不同处理方法的效果。定性方法较适合于研究那些处于变化和发展过程的事物，以观察事物的变化和发展在观察对象身上所发生的影响。这些影响可以是预期的，也可以不是观察者所能预料的。因此定性方法是一种面向过程的动态观察，不但能够反映预期的结果，而且也要捕捉那些非预期的后果，特别随着环境的变化而变化的因素。在语言研究方面，定性方法最适合于观察在各种不同的环境里，不同的人群的语言使用的情况。

但是，语言系统在动态发展中往往又是相对稳定的，这样我们才能对它进行客观的描述。Saussure认为语言像下棋一样，它是不断地变化的。但在每一个阶段，棋盘上的每一个棋子和其他的棋子都保持着某种特定关系，往一个方向的移动为另一方向的移动所平衡，因此它

又是相对稳定的，或称为动态的平衡。

6.1.7. 抽样

表面看来，调查者在现场调查中尽量如实记录现场的活动，无需抽样。实际上，我们很难做到每一件事物都作详细的观察，调查者看到的仅是活动的一部分，所以他所观察的其实是他所能看到的所有事物的一个样本。有时观察者有可能在所观察的事物中作一些选择，这就需要采用一些适合于现场观察的抽样方法[①]。不像实验方法所用的控制概率抽样法，定性研究所采用的是有目的的抽样（purposive sampling）：选择那些你认为最有助于了解你的研究对象的样本来进行观察。

一种方法是定额抽样（quota sample）。如果所观察的事物、人群、过程有明显的界线，就可以采用定额抽样的方法，使不同类型的单位都能包括到样本里。定额抽样要求我们对总体有所了解。Gallup 在 1936 年根据他所掌握的美国不同阶层的收入水平，决定每种样本的数目，成功预测了罗斯福在大选中的胜利。在 1940 和 1944 年也同样取得成功。但是在 1948 年的预测中，他却失败了，主要的原因是在战争中美国人口大量从农村往城市转移，使他根据 1940 人口调查的数据所选的样本缺乏代表性。

第二种方法是滚雪球抽样（snowball sample）。例如我们想了解吸毒者怎样染上毒瘾的型式，可以先就便找几个吸毒者来了解，然后根据他们提出的线索，逐步扩大范围，由近而远，再归纳出染上毒瘾的型式。

第三种方法是异常个案（deviant cases）。有时，我们可以通过研究那些不符合规则的型式的异常个案来加深我们对态度和行为的型式的了解。例如大家对某一个演出节目反应强烈，起劲地鼓掌，但是也有几个人不鼓掌，不妨访问一下他们，从另一个角度去了解群众反应，我们的认识就能更为全面。

在通过抽样来进行现场调查时，应该注意两个方面的问题：

1. 可供观察的全部情景在多大程度上代表了你所要进行描述和解释的现象的一般的类型？例如你要了解某种语言现象，决定向三种类型的资料提供人调查，这三种类型的人能否代表总体？

2. 你在全部情景中所进行的实际观察能否代表了所有可能的观察？你所调查的对象是否真正代表了这三种类型的资料提供人？

虽然我们在定性研究中甚少采用控制的概率抽样，样本的代表性和结论的概括性仍然有密切的逻辑联系。在我国语言学研究中，一些文章往往采用了定性的方法，从一定理论框架或观点出发，然后收集符合它们的资料，便作出理论性的概括。因为这些资料缺乏代表性，或有偏颇性（有意无意地排斥不符合自己观点的资料），所以结论就不够科学。所以有目的的抽样仍然要符合抽样原则，而不是各取所需。

---

① 在本书里，我们在不同的章节都会谈到抽样（sampling），这里是从现场调查的角度来谈有目的的抽样，更详尽的讨论见 8.2，抽样数目的决定见 11.8。

6.2. 进行定性研究的程序

如上所述，定性研究是从一般到特殊的研究，所以有人把它描写成一个倒金字塔或漏斗。另外还有人把它比作一个螺旋体，因为研究者收集从一般到特殊的数据时，往往要经历过一些观察和分析的重复性的周期：随着研究的进展，每一个连续性的分析阶段都会把研究者的注意力导向不同的观察面。所以定性研究一般来说都是比较开放性的，往往取决于特定的研究环境，没有什么事先规定的严格程序，下面所谈的仅是一般的做法。

6.2.1. 定义所要描写的现象

定性研究采取综合方法，这意味着观察的范围必须在某个阶段缩小。在研究的早期，观察者可以不带任何框框，全面地接触现象，对它们采取开放性的态度。但是到了后期，就必须缩小观察的范围。各种行为都是有层次的，大单位套小单位，因此我们第一步是定义所要描写的现象，也就是说，必须决定在哪一个层面上进行观察。例如在描述一种语言时，语音、形态、词、词组、短语、句子、段落、篇章都是可以描述的层面，但是层次越高，分析就越困难，因为层次高的单位包括很多子单位，子单位本身又包括很多子单位。而且越高的单位和使用的环境关系越密切。Hymes（1972）指出分析言语行为时，应该区分三个层面：

1. 言语环境（speech situation）
2. 言语事件（speech event）
3. 言语行为（speech acts）

言语行为是最小的一个分析单位，如问候、道歉、表扬、介绍、提问等等。言语行为根据某些社会规范组成一个连续的系列，就是言语事件，如交谈、访问、和售货员对话、打电话查询等等。言语环境是说话发生的上下文——即与言语行为相联系的典型环境，例如生日聚会、家庭晚餐、拍卖、看足球赛等等。我们以一个校友聚会为例，说明这三个层面的关系。聚会本身是一个言语环境：它有始有终，通常在一天内完成；参加者限于老同学和他们的配偶。在这个言语环境里，会有一连串的言语事件：一群人可能在回忆他们过去的老师或课堂里的恶作剧；另一群人则在谈论他们毕业后的事业发展等等。在每个言语事件里开个玩笑，谈点个人经历是言语行为。

对所描写的现象进行定义也可以说是建立一个概念框架，这个概念框架规定了我们所要研究的维度——重要的因素和变量，而且预测它们之间的关系。然后在这个框架里，形成我们的研究课题。

6.2.2. 使用定性方法收集数据

定性研究所收集的数据主要是词语，而不是数字。词语能够直接地、具体地、生动地说

明现象，说服力较强，连采取定量研究方法的人也越来越认识到定性数据的重要性，因而注意收集定性的数据。定性研究使用各种手段收集数据，甚至在同一研究中也使用不同的方法，以窥全貌。常用的方法有观察、录音、问卷、访问、个案、现场记录等等。由于收集数据的来源和方法多种多样，定性研究比单一的实验或测试能够获得更多的信息和启发。

数据的收集根据概念框架所规定的范围来进行，所以从一开始就按照分类系统的原则，制订一些严密的整理数据的表格，使数据得到妥善的保存和能够迅速提取。一般人以为收集数据和分析数据是两个不同的阶段，但是实际的情况是收集和分析往往不容易分开。我们在收集数据的同时，就要开始分析数据，看看所获得的数据符不符合概念框架的要求，并生成新的收集数据的方法，克服观察的盲点。所以从一开始，就应把数据的收集和分析结合起来。整理数据的表格可以起到收集和分析相结合的作用。

定性数据主要是词语，但是词语也有它的问题，主要是它比数字复杂，而且一词多义，离开了上下文就没有意义，如“这个”，“它”是要看前指词才能断定其所指的，汉语的 shizi 可以是“狮子”，也可以是“虱子”。英语的“board”可以是“木版”，也可以是“委员会”。数字则没有那么含糊，而且处理起来方便。要保持词语和数字的优点，就要对词语进行编码。在社会科学和人文科学研究中应用计算机技术已经越来越普遍，计算机可以把大量的信息保存成为文件和数据库，对资料的整理、重新组合和提取提供了很大的方便。但是要发挥计算机的作用，我们必须对信息进行编码。在上面我们已经谈到建立分类系统的必要性，编码无非就是把分类系统代码化。

对资料进行编码可以有两种方法。一种是根据研究的需要已经建立了一个相对地完善的编码方案，如中国社会科学院语言研究所所公布的《方言调查字表》，我们需要做的无非是把编码范畴定得更为细致和具体，然后训练编码人掌握这些范畴，调查后还要检查他们是否正确编码等等。但是如果我们在一开始时对数据的编码并没有把握，我们还不知道数据代表什么变量，那就要采取另一种方法，先拿 50－100 张调查表进行分析，根据反应结果建立一个初步的编码范畴，然后再来对别的调查表进行编码。编码范畴必须是详尽的，而且是互相排斥的。也就是说，每一项信息只能放在一个范畴里面。还有一种介乎两者之间的方法，就是先提出一个一般的编码范畴，然后再逐步深化，Bogdan & Biklen（1982）认为任何研究都应该处理下面几个从微观到宏观的关系：

1. 背景/环境：对周围环境的一般的信息。
2. 情境的定义：人们是怎样对所研究背景进行定义的。
3. 观点：思维和取向的方式。
4. 对人和物的思维方式：比 3 更为详细。
5. 过程：先后次序，流程，随着时间而发生的变化。
6. 活动：经常性的行为。
7. 事件：特定的活动。
8. 策略：完成事物的方式。
9. 关系和社会结构：非正式规定的型式。
10. 方法：与研究有关的问题。

社会语言学家在语言调查中还发展了一些收集数据的新方法，叫做提取法（elicitation

method)，见 8. 3. 2. 3。

怎样提高数据和数据收集过程的质量是每一个研究者都关心的问题。首先是数据本身的质量，在研究过程中，往往有些外部因素影响我们对数据的解释，使我们的观察失效，例如总体的特征是否有足够的代表性？样本或资料提供人的选择是否客观可靠？分类系统中所规定的变量是否清楚明白？研究环境会不会产生特殊的效应？研究者或调查人会不会对资料提供人产生特殊的影响？这是外部效度的问题。如果我们所得到的结果不能延伸或应用到观察环境以外的环境，这就是外部无效。有时候我们的研究方案本身也足以影响结果，这就是内部效度的问题，例如样本或资料提供人是否随机的？样本和资料提供人是否足够大？时间因素会不会影响数据的收集？所采用的调查工具是否足够敏感？

其次是数据收集过程是否前后一致和准确无误，这是信度的问题。例如作语言调查，特别是遇到一些不太明显的语言现象时，如果没有一张调查的细目表，调查人就很容易有意或无意地出现偏颇，没有客观而全面地记录语言现象。这就需要提高评估员间信度（inter-rater reliability），看看不同的观察者对数据的处理是否一致。还有测试-重复测试信度（test-retest reliability），把调查程序重复做一次，看是否得到同样的结果。有时，调查者使用不同方式的数据调查程序，那就必须考虑平行调查卷信度（parallel form reliability ），看看这两种方式是否能收集同类数据。如果一份问卷或考卷有若干个了解同一内容的独立项目，那就需要看这些项目是否都能抽取同样的信息，这就是内部一致性（internal consistency reliability）。由此可见，不同的数据收集过程可能需要不同的信度检验。在定量研究中，信度是一个从 0. 00 到 1. 00 的指数。在定性研究中，不一定能够取得一个量化的指数，但其原理仍然适用。[1]

6.2.3. 在数据中寻找型式

定性研究并非从特定的研究课题或假设出发，所收集的都是原始数据，并没有进行筛选。研究人员必须从数据中寻找反复出现的型式。这也可以称为型式编码。型式编码是有解释力的编码，它可以找出主题、型式或问题。它把众多的材料归纳成更有意义的分析单位，因此又可叫做元语码（meta-codes）。Harris（1951）在他的《结构语言学方法》里指出，“描写语言学是一种特定的考察领域，它并非研究整个言语活动的，而只是研究言语某些特征的规律性（regularities）。这些规律性存在于所考察的言语特征的分布关系之中，即这些特征在话语相互之间的出现频率。”由此可见，型式，也就是 Harris 所说的规律性，主要是指要素之间的某些稳定的相互关系。在型式的基础上，研究者就可进一步形成假设，甚至提出一些解释调查结果的模型。例如我们要研究人们在会话中是怎样轮流说话的，第一步是把会话录音或录像，然后是对它们反复观看，看有多少轮流说话的型式，最后才形成假设，提出模型。

Miles，M. & Huberman，M.（1984）认为型式编码有 4 个重要的功能：

1. 它把大量数据归纳成更小的分析单位。
2. 它使研究者在收集数据时就开始进行分析，使以后的数据收集得更为集中。
3. 它帮助研究者建立一个认知图式，以了解局部发生的事情。

---

[1] 在 10.7.1 和 10.7.2 里还会分别讨论信度和效度以及它们的估算方法。

4. 当几个研究者在进行个案研究时，它把共同的主题和因果关系显露出来，为跨场合的研究奠定基础。

6.2.4. 回到数据中去收集更多的数据以支持初步的结论

一旦型式找了出来，定性研究者就需要检验结论的效度。由于我们采取了各种方法去收集数据，就有可能使用三角验证（triangulation）的方法来检验效度，就是说，看看不同的数据源能否产生同样的型式。使用这种检验方法可以增加结论的效度。

一般来说，导致定性研究失效的原因有两种：

1. 研究不是在自然环境中进行或所收集的数据并没有代表性。这就是说，有些东西在数据收集的过程中对所观察的行为作了歪曲。解决的办法是重复该项研究，看第二次所收集的数据能否归纳出同样的型式。重复研究可以检查在第一次观察时有无使观察行为异于正常的情况发生。

2. 没有控制可能在定性研究的数据解释中出现的主观因素。三角验证方法可以控制主观性，例如在研究会话中的轮流说话，我们可以在录音或录像的同时还作现场记录，互相印证，还可以让别人来分析结果，看是否得到同样的结论。这里说明，定性研究的数据必须具有可提取性（别的研究者可以看到原始数据）和可证实性（别的研究者可以独立验证结果）。

6.2.5. 研究过程的循环往复

经过第一阶段的数据分析后，可能我们还要对研究范围重新定义，缩小研究范围，这是一个使焦点更为集中的过滤过程。当研究的现象越来越集中时，研究者就必须再作一次数据收集和分析的循环。例如我们从数据中发现男人比女人在轮流说话中更为主动，那就需要缩小范围。收集更多的数据来验证这个型式。

6.3. 定性研究、描述性研究和实验性研究

在上述讨论中，我们常常拿定性研究和实验性研究来比较，因为这两种研究有较大的差异。其实在它们之间，我们还可以区别出一种间乎两者之间的研究，这就是描述性研究（descriptive research）。

描述性研究和定性研究有共同点，也有差异点：例如（1）它们不像实验方法那样控制和操纵变量，都强调自然观察，不干预所调查的现象；但定性研究主张在调查前不带任何框框，不提出什么假设，而描述性研究则主张可以根据现存的数据或现象事先提出假设，作为考察的基础。（2）它们都强调综合的系统方法，主张全面地观察问题，但描述性研究不排除使用分析方法，以集中观察一些特殊现象。换句话说，描述性研究者从一般的问题出发，但又将焦点聚于一些特殊的问题。（3）它们都主张归纳式，但描述性研究不排除使用演绎的方法。（4）定性研究较少采用定量手段，但描述性研究则在可能的范围内使用定量手段。除了上述的异同点外，进行描述性研究的程序和定性研究的程序基本上相同。

由此看来，描述性研究脱胎自定性研究，但在可能范围内也采取了一些定量研究的手段。应该说，很多人类语言学、描写语言学、结构语言学、社会语言学，乃至统计语言学和语料语言学的研究都采用了描述性研究的手段，所以我们把中篇叫做语言描写方法，它以描述性研究为主，但也包括了纯粹的定性研究和一些定量研究。读者从以后的讨论中可看得很清楚。

定性研究、描述性研究和实验性研究的区别可以见于下图：

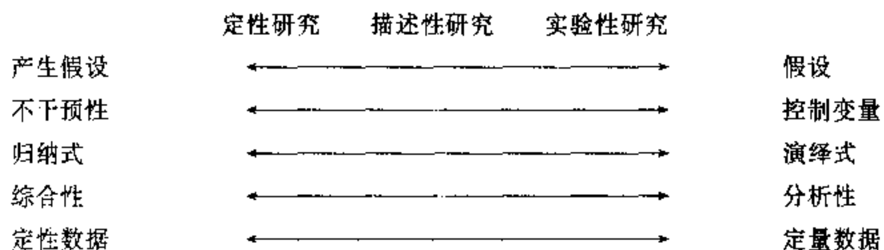


图 6.5

7. 语言描写方法

7.1. 现代语言学的诞生

19 世纪末期自然科学的发展使一些语言学家（如 Schleicher）从生物学进化论的角度去看待语言，诱发了历史对比语言学的产生。早在 1786 年，英人 Jones 发表了一篇文章，指出梵语和希腊语、拉丁语有很多相似之处，由此导致了一系列的研究，终于确立了印欧语谱系。历史比较法比较方言或亲属语言之间的异同，找出它们相互的语音对应关系，以确定语言间的亲属关系，并进而构拟原始“共同语”。这种方法和生物学家用来研究生物进化的方法很相似，例如生物学家通过比较各种类人猿，发现它们都有五个能抓东西的手指，其中有一个大拇指，而且都有指甲，推断这些动物有共同的祖先。美国的 Bloomfield 在 20 年代也试图使用比较法来研究北美的一群 Algonkian 语，他根据自己在北美五大湖区所收集的四种 Algonkian 语的资料，构拟了一种原始 Algonkian 语。后来的研究者发现 Bloomfield 没有调查的其他几种 Algonkian 语也很符合原始 Algonkian 语的轮廓，终于建立了原始 Algonkian 谱系。这种历史比较法基本上采用了定性研究方法，先是调查研究，收集资料，然后从资料中寻找型式，故法国语言学家 Meillet（1924）认为，每一种古代的大“共同语”应该表现一种文明类型，所以世界上的大部分语言看起来都是由少数几种共同语演变出来的。

作为现代语言学诞生的标志是描写语言学或结构语言学的出现，这和现代科学的发展也不无关系。Winograd（1983）在讨论到语言学的进化时，把历史比较语言学比作生物学，而把结构语言学比作化学。结构语言学在始于欧洲（de Saussure），但在美国（Bloomfield）得到发展，它把注意力转移到共时语言的个别语言的描述，而且受美国心理学中的行为主义的影响甚深。行为主义者认为从心理过程去解释人类行为是不科学的，因为心理活动不可捉摸，我们只能依靠客观的观察来分析人类行为，包括语言行为。这一点和人类语言学重视现场调查不谋而合。语言学家只能研究在自然环境中产生的语料。化学家采用实证科学的办法去研究物质结构的做法对语言结构的研究给予很大的启发。化学家通过实验去决定一个复杂的物体由什么分子组成，并从基本元素的角度去分析这些分子。化学的最大的成就在于找到一些基本元素，这些元素通过不同的组合，构成了我们的大千世界。门捷列夫的“化学元素周期表”是根据元素性质的周期性变化而排成的，展示了各种元素性质递变的规律。他甚至根据这些规律预测新元素的的发现，并推断新元素的性质。通过要素的排列来发现要素之间的规律对语言学研究产生了重大影响，例如根据基本音素的排列而制定了国际音标。

美国描写语言学所采用的方法为现代语言学的数据收集和分析奠定了基础，60 年代以后，Chomsky 的转换生成语法打破了结构语言学的一统天下，出现很多语言学流派，尽管它们的理论模型很不一样，但是仍然在不同程度上继承描写语言学的方法。

7.2. 语言描写的基本概念和模式

Robin (1989) 在他的《普通语言学》里指出，语言学和语言研究在科学中占有特殊的地位，因为语言学家既是语言系统或具体语言的观察者，又起码是一种语言——他的母语的产生者和评估者。这意味着语言学家既可以采取一种数据的"外部的"观察者的立场，又可以采取一种"内部的"分析家的立场。从"外部的"观点看来，语言学家和别的科学家（如物理学家、化学家、生物学家）一样对待他的资料：观察、分类、寻找内部规律性、提出假设来检验更多的数据。别的观察者也可以在相同的基础上占有他所处理的所有数据。Bloomfield和他的追随者，采取的是这种"外部的"立场。

从"内部的"观点看来，科学家是在观察自己，不但要了解他所说和所写的牵涉到一些什么东西，甚至要了解他的大脑是如何运作的，为什么他能产生和了解无限的母语句子。语言学家是以说话人/听话人的身份去分析他自己使用的语言。别的操母语者也同样可以占有他使用的数据；但是就每一个人而言，这些数据都是个人现象，不能直接和公开观察。Chomsky和他的追随者采取的是这种"内部的"立场。根据同样的思路，Kibrik (1977) 列举了描写语言学的三个关键的概念：

1. 考察的主题（语言或一部分语言）
2. 考察的对象（书面文本或录音数据）
3. 考察的成果。考察主题的模型，即语法。这是广义的语法，包括语言理论和具体的语言变异模型。

这三个概念其实表明了一种调查者/数据/理论的三角关系：

首先是调查者和数据的关系。调查者可以采取不同的方法来进行一项共时的语言描写。"描写"的概念很广泛，它包括描写一种调查者一无所知的语言或描写调查者自己的母语。如果调查的是一种调查者不懂的语言，那就要依靠资料提供人；如果调查的是自己母语的用法，那就可以依靠资料提供人或自己的母语的直觉（即用内省法）。所以假定输出的总是（不同程度的）理想化的语言模型，我们就可以从使用数据的类型来考察调查者和数据的关系。

其次是数据和理论模型的关系。根据数据而建立的语法模型也是多种多样的：可以是"核心"语法的某些方面，如英语的名词短语结构；可以是某种语言的各个语法范畴；可以是一个言语社区的某些成员系统地使用的某些变量（如 Labov 在纽约市调查的 5 个社会语言学变量)。这些理论模型都是长期的社会的、数学的、语言的抽象化的结果，都可以说是理想化的语言模型，不管它们所采用的方法、理论目标和假设有些什么不同，它们和数据的关系都是间接的。这说明一条大家都接受的语言学原理：语言系统是一个不能进行观察的，抽象的，而具体的话语却是可以观察的；这些话语体现了掌握了该语言系统的人的语言能力。

Kibrik 把这种调查者/数据/理论的关系表现为图 7.1：

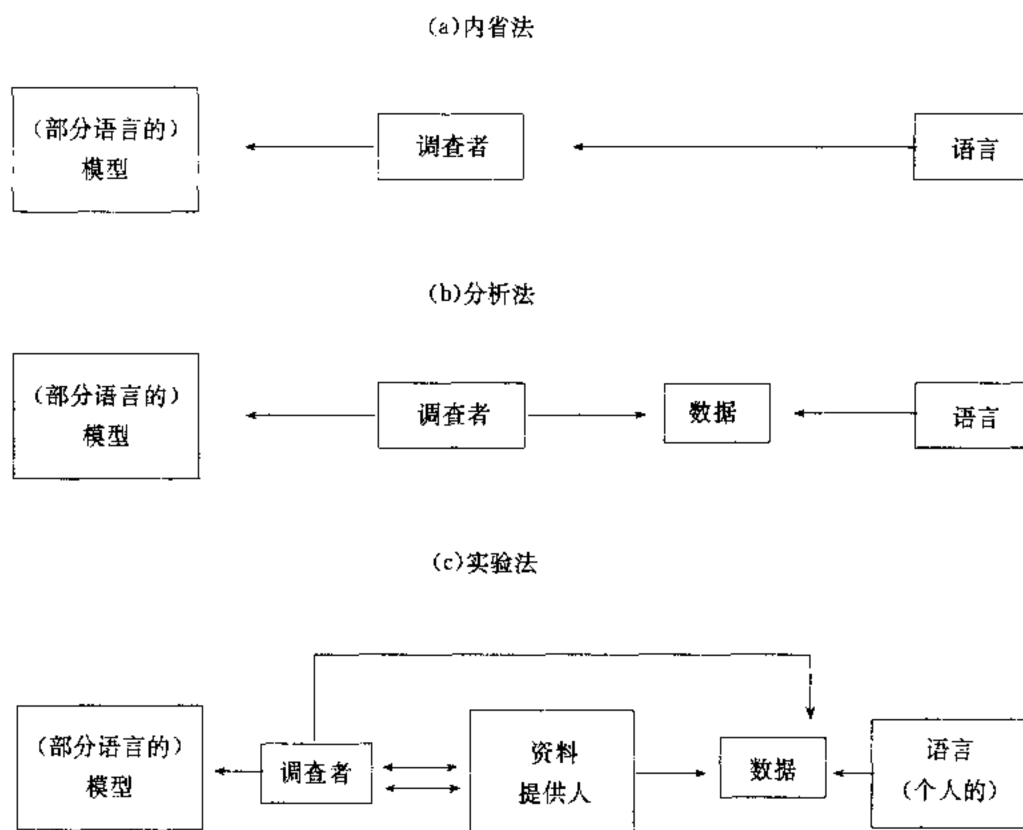


图 7.1

图 7.1 (a) 表示的是调查者通过自己的语言能力直接接触目标语，这种语言的描写方法是内省式的自我观察（有时可用别人的内省结果来检查），所以它没有调查的对象，故图中没有“数据”。这种方法不能用来研究调查者所不了解的语言或语言变体。因此，除非语言学家本人的母语是非标准的方言或没有文字的语言，否则他只能用这种方法来研究规范化的语言。所以，要描述非标准的方言时，不能把语言直觉作为基本的方法，而只能把它作为一种帮助我们聚焦的手段。虽然内省法也能了解到语言的结构和组织的许多事实，但是它的应用范围不很广。理论语言学家较多采用这种方法。

图 7.1 (b) 所表示的方法也牵涉到调查者本人必须了解目标语，但他根据独立收集的数据来作出自己的理论概括，而不是依赖自己的语言直觉。一些语篇分析和文体学的研究者往往采用这样的方法，因为一些关于语篇分析和文体学的理论概括是难以凭语言直觉作出。但是我们能否凭这种方法，只通过语料分析就可以了解到我们一无所知的语言结构呢?Chomsky 对这一点攻击甚烈，但事实上光用这种方法的情况是很少的。一般的情况是，除了依赖语料分析外，语言学家还采用了别的方法，(a) 或 (c)。例如 Quirk 等人 (1972) 所写英语语法既使用了分析法，又依赖了内省法来提取语言结构的信息。

图 7.1 (c) 所表示的是 Kibrik 所谓的“实验法”，应该指出的是他说的方法实际上是我们在以上谈到的现场调查的方法，并非我们在下篇所讨论的实验方法。这种方法使用独立于内省观察之外的数据，往往是使用一个操母语的资料提供人来提供关于目标语或语言变体的语

言事实。多数的社会语言学和方言调查把它和（a）和（b）结合起来使用。如果考察的是一种调查者所不知道的语言，往往是（b）和（c）一起使用。

7.3 语言描写的目标——抽象化

如上所述，语言描写本身不是目的，其目标是建立理想化的语言理论模型。所以语言分析和其他学科一样，都要在不同层面上对它的各种要素以及它们相互关系进行抽象化，这些要素具有标志着它们相互关系的各种恒量、范畴和规则，我们研究和解释语言现象时会经常使用。在任何一个层面上，抽象化可有不同程度的概括，例如从语法上看，“地图”（从某些话语中抽象出来声音）、“名词”、“词语”可以说是不同程度的抽象化。抽象化的程度越高，它所包含的东西就越多：“名词”不但包含“地图”，还包含“书”、“白菜”、“工人”、“铁”等等。从语音上看，“辅音/b/”、“爆破辅音”、“辅音”也是不同程度的抽象化。

抽象化的东西都可以间接地体现为口语的或书面语的形式，而且一个抽象的要素往往还可以有几个相应的形式，例如在英语里，名词的复数在口语里可以是/-s/，/-z/，/-iz/，/ˈɔksn/（oxen），甚至/men/（men）。同样的，我们可对拉丁语、德语的格的变化，对很多语法范畴进行抽象化。抽象出来的要素和实际形式之间的间接的、复杂的关系在转换生成语法中更为明显。

Robin 指出，抽象化在语言形式的分析和描写中的地位是一个争论颇多的话题，但下面三点在不同的程度上都得到多数人的支持：

1. 如果分析正确无误，抽象出来的恒量应该以某种形式存在于所分析的语言材料之中。这种观点意味着语言结构是独立于分析家之外的客观实体，而分析的任务则是发现该结构的一些内部型式。语言学家的结论是否正确取决于结论是否符合事先存在的结构的实际。

2. 操母语者也能作出和语言学家的抽象相同的抽象，但他们的抽象是无意识的：它是某些内在心理结构的产物，是他们在儿童阶段习得母语的结果，是说话人的心灵和大脑的内容的一个部分。这种观点意味着语言学家根据观察所作出的语言分析在某些方面和操母语者的心灵和大脑中的语言机制是一致的。de Saussaure 的“语言”（langue）和“言语”（parole）的区分，Chomsky 的“语言能力”和“语言运用”的区分，都和这一点相吻合。de Saussure 认为，语言是一种语言的词汇、语法、语音组织，从小就作为一个言语社区的集体成果植根在操母语者心中或脑中。说话人在说话时只能在这个语言组织的范围内运作，他所实际说出来的是言语。Chomsky 所说的语言能力指操母语者对语言的直觉的知识，而语言运用则是他在实际使用语言时所做的事情。

3. 语言学家所作出的抽象无非是他的科学术语的一部分，用来陈述规则和预测话语的形式，这是无可厚非的。这些科学术语和操母语者的心灵或大脑状态不排除有某些对应关系，但两者之间并无直接关系。

这三种观点对语言学研究来说，都有一定的效度，多数语言学家或多或少地持有这种或那种观点。一般来说，持“内部的”观点的语言学家采取第（2）个立场，而持“外部的”观

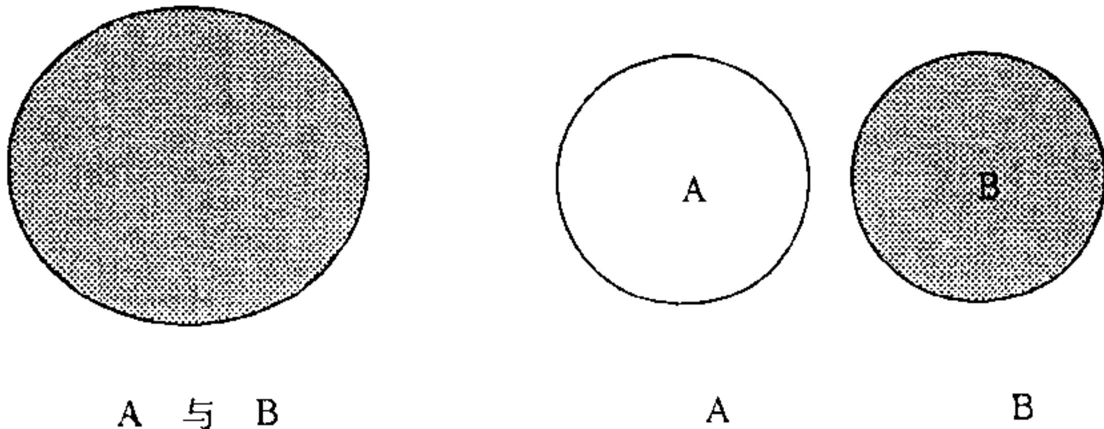
点的语言学家赞成（1）或（3）。

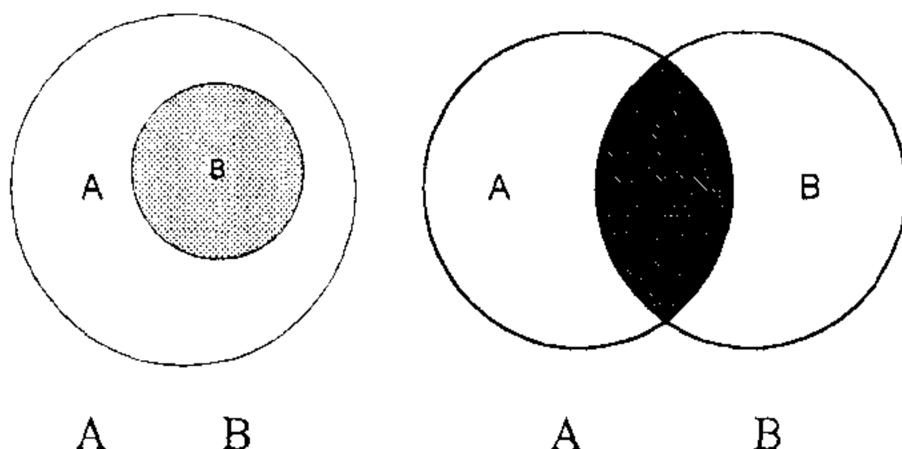
7.4. 语言要素的分布

Harris，Z.（1951）在讨论到结构语言学的方法时，首先谈到的是关联准则（criterion of relevance），即分布。“描写语言学是一种特定的考察领域，它并非研究整个言语活动的，而只是研究言语某些特征的规律性。这些规律性存在于所考察的言语的特征的分布关系之中，也就是在话语中这些特征相互之间的出现频率。我们可能研究言语的各个部分或特征之间的各种关系，例如声音或意义的相同性（或其他关系）或语言历史中的亲属关系。”

Harris 进一步说明分布的概念：假定言语的音素表征是一对一的关系，这并不意味着如果某一声音 x 和音素 Y 有联系，只要给出 Y，我们就把它和原来的那一个声音 x 联系起来。一对一的对应关系只意味着，如果某一特定声音 x 在某一特定的位置和音素 Y（或符号 Y）有联系，只要给出音素 Y，我们就在指定的位置和某些可以替换原来的 x 的声音的 x'，x"（即在分布上和 x 相同）联系起来。符号 Y 在指定的位置可代替任何可与 x'，x"互换的声音。所以 Haugen（1950）指出，“分布是区别语音学的音位学钥匙和区别语义学的形态学钥匙。这是数学方法和语言描写中的共同因素。……（分布）的根本的方法，即替换，所指向的首先是发现分布。这种技巧跟自然科学的控制试验有些相同。……替换就是一个明显的成分实际上替代了另一个明显的成分。”

按照 Bloomfield 的定义，句子是一个独立的单位，不能凭借任何语法结构而包含在任何更大的语言形式里面。除了句子以外，每个语言单位都或多或少地局限于它所出现的上下文。如果两个或更多的单位发生在相同范围的上下文，它们就可以说是等值分布，即具有相同的分布（图 7.2a）；如果它们没有相同的上下文，它们就是互补分布（图 7.2b）；如果一个单位的分布包含了另一个单位的分布，那就是包含分布（图 7.2c），例如 x 出现在所有 y 出现的上下文，但是某些上下文只有 y 出现，而没有 x 出现；如果两个单位的分布交叉，那就是重叠分布（图 7.2d）。例如有些上下文是 x 和 y 都出现的，但是它们各自又出现在自己的上下文里。





A　　B　　　　A　　B

图 7.2

应该指出的是，“分布”既和一个语言单位所出现的上下文有关，这个单位的出现就必然受到一定的制约，而这种制约是可以系统地表述的。例如英语中的辅音/l/和/r/是部分等值的，它们都能出现在一些除了这个音以外的语音上相同的词里，如：light：right，lamb：ram，blaze：braise，climb：crime。但是这是有限制的，不是所有的上下文都如此，例如 slip 没有相对应的 srip ，trip 没有相对应的 tlip，blend 没有相对应的 brend，brick 没有相对应的 blick。在这些不可能出现的词里，srip，tlip 和 brend，blick 又是有区别的。前者可根据英语的语音结构的一些规则来排除的（即作出系统的表述），英语中没有/tl/和/sr/那样的辅音丛；而后者的辅音丛/br/和/bl/是不能作出系统表述的，因为它们可出现在 blink：brink，blessed：breast 的上下文里。因此 brend 和 blick 在英语中是可以接受的，只不过英语刚好没有这两个词。

由此看见，每个语言单位都有一个对立功能（contrastive，de Saussure 称为 opposition）和一个组合功能。两个单位必须起码在分布上是部分等值的，才能对立。出现在不能对立的上下文里的两个单位就可以说是自由变异（free variation），例如英语 leap 和 get 出现在多数上下文里时是对立的（试比较 beat：bet），但在 economics 这个词的发音时，却是自由变异的，它的第一个元音既可发成/i/，也可发成/e/。自由变异是在上下文里的功能等值，它和等值分布（出现在相同范围里的上下文）是不一样的。两个对立的因素在相同的上下文里互相替换的结果是产生了不同的词；自由变异则不会这样。

由此看来，分布意味着要素之间构成一种互相联系系统，它们各司其职，但又相互联系，成为一个整体。Robin 曾经把这种关系比喻为一个乐队，乐队里的每一个成员在整个乐队和乐器小组里都有其自身的职责，每一个人都处于相对于其他人的一个位置，完成自己的功能。增加或减少一个队员都会影响演出的质量。这和音乐会的听众不同，听众仅是一个简单的集合，是男是女，多几个少几个，对听众容量的影响都不太大。

7.4.1. 组合关系和聚合关系

语言要素的关系有两种：一种是组合关系（syntagmatic），另一种是聚合关系（paradigmatic）。

组合关系指在某一个层面上各个要素之间的关系，这些关系把要素组成系列的组织，是一种横向的关系。例如英语中的短语 take care 可以在不同层面上有不同的组合关系：表示为音标的/'teik'kεə/是一个层面，更加抽象的语音表征 CVVC CVV（C＝辅音，V＝元音）是另一个层面，语法排列的动词＋名词又是一个层面。它们都可以说是不同层面上的组合关系。这些组合关系都是指向口头的或书面的话语，它们可以说是主要的维度。

聚合关系指处于结构中特定位置的可比因素之间的关系，因而是一种纵向的关系。例如：

首辅音　take　/teik/

　　　　m　/m/

　　　　b　/b/

动词后的名词 take care

　　　　pains

　　　　thought

　　　　counsel

应该说明的是（1）组合关系和聚合关系是互相依存的，换句话说，一个语言单位和别的语言单位既有组合关系，又有聚合关系。没有这两种关系，一个语言单位就不成为语言单位。这也就是说，每一个语言单位在一个多种关系的系统中都有它的位置。(2)“组合”并非都含有“次序”的意思。组合关系并不意味着语言单位的次序必须是线形排列。Lyons，J.（1968）曾以汉语的声调为例，说明一个词的声韵和声调是难以分出谁先谁后的。说话虽然有时间次序，但是这个次序是否和语言结构有关，主要是决定于语言单位的组合关系和聚合关系，并非决定于谁先谁后。(3) 在语言学的讨论中，“结构”和“系统”常常是可以替换使用的，但有些语言学家（如 Firth）则主张作出一些区别，(语言）结构可用来表示具有组合关系的要素，而（语言）系统则表示具有聚合关系的要素。在结构中有子结构，在系统中也可有子系统。

7.4.2. 形式化

分布分析不但对描写语言学，而且对整个语言学的发展都起了重要的作用。Lyons 举了一个简单的例子来说明它对语言形式化的影响：假定我们所分析的语料只有 17 个“句子”：ab，ar，pr，qab，dpb，aca，pca，pcp，qar，daca，qaca，dacp，dacqa，dacdp，qpcda，acqp，acdp。每一个字母代表一个词。我们可见 a 和 p 具有某些相同的环境（如：-r，pc-，dac ），而 b 和 r（如：a-，qa-），d 和 q（如 dac-a、-aca，ac-p）也一样。只有 c 的分布比较独特（如：a-a，p-p，qa-a，da-a，da-p 等等），没有别的“词”出现在 c 所出现的环境。如果我们把 a 和 p 归为 X 类，并把 X 代入 a 和 p 出现的地方，就有 Xb，Xr（ar，pr），qXb，dXb，XcX（aca，pca，pcp），qXr，dXcX，qXcX（daca，dacp），dXcqX，dXcdX，qXcdX，XcqX，XcdX。如果我们把 b 和 r 归为 Y 类，把 d 和 q 归为 Z 类，并把 Y 代入 b 和 r 出现的地方，把 Z 代入 d 和 q 出现的地方，就有（1）XY（Xb，Xr）；(2) ZXY（qXb，qXr，dXb）；(3) XcX；(4) ZXcX（qXcX，dXcX）；(5) ZXcZX（dXcqX，dXcdX，qXcdX）；(6) XcZX（XcqX，XcdX）。这样我们就能用 6 条结构公式来说明语料库的 17 个“句子”，并规定哪些词的次序是可以接受的。

这 6 条规则中的每一条分别描写了一种句子类型，而且说明语料库的所有 17 个句子均可

接受（即符合语法）。其实，它们可说明的不只这 17 个句子，而是 48 个句子。第一条规则有两个要素（X 和 Y），X 包括了 a 和 p，Y 里包括有 b 和 r，所以可能的组合是四种（2×2），这可以表示为一条公式：

$N = p_1 \times p_2 \times p_3 \cdots\cdots p_m$（N 为高一层面的单位数目，m 为低一层面的要素的组合对立的位置数目，p1 为第一个位置的聚合对立的要素的数目等等）。

1. 为 2×2＝4
2. 为 2×2×2 ＝8
3. 为 2×1×2 ＝4
4. 为 2×2×1×2 ＝8
5. 为 2×2×1×2×2 ＝16
6. 为 2×1×2×2 ＝8

所以这个语法所能描写的语言合共有 48 个句子，有 31 个句子没有出现，但都是可以接受的。这个语法生成了所有的句子，而且给予每一个句子一种结构上的描述：pr 属于句子结构 XY，pcda 属于句子结构 XcZX，诸如此类。

7.4.3. 一个实例

我们还可以举一本具体的结构主义语法为例来说明分布原理的应用，Fries，C.（1952）的《英语结构》提出三个句子的框架：

(A) The concert was good (always)
(B) The clerk remembered the tax (suddenly)
(C) The team went there

然后根据聚合关系的原则，把处于同一位置的词通通归为一类，以（A）为例：

The *concert* was good.
food
coffee
taste
container
difference
privacy
family
……

能够放进这个位置的就叫做一类词（相当于传统语法的各类名词和动名词），因为这类词可以

加“s”而变为复数，而动词还有个时态的问题，所以框架（A）还可以补充为：

(The) ________ is/was good.

……s are/were good.

一类词还可以出现在框架（B）和（C）里：

(B)

The *clerk* remembered the *tax*.

husband	*food.*
supervisor	*coffee.*
woman	*container.*
	difference.
	family.

(C)

The *team* went there.

husband

woman

supervisor

Fries 一共提出 4 类词，第二类词为各类动词，第三类词为各类形容词和分词，第四类词为各类副词，例如：

(A)

一类词	二类词	
(The) ________	*is/was*	*good.*
________ *s*	*are/were*	*good.*
	seems/seemed	
	seem	
	sounds/sounded	
	feels/felt	
	feel	
	becomes/became	
	become	

(B)

一类词	二类词	一类词
(The) ________	*remembered*	(the) ________.
________ *s*	*wanted*	________ *s.*
	saw	
	discussed	
	suggested	

understood
signed
preferred
stopped
straightened

(C)

一类词	二类词
(The) ________	*went* there.
________ s	*came*
	ran
	started
	moved
	walked
	lived
	worked
	met
	talked

同样的，第三类和第四类词也可放进（A）(B)（C）三种框架里。另外 Fries 还把功能词分为 15 组，如 A 组包括了冠词、代词、数词、名词所有格，等等。限于篇幅，不再列举。

和其他的结构主义者一样，Fries 有意回避了传统语法中的各种词类的提法，而索性用简单的分类，因为（1）传统语法对各种词类的定义含混不清，有的从意义上定义，有的从功能上定义；（2）英语中不少词有形态特征，可根据形态决定词类，如 goodness，departure 为一类词，soften，befriend 为二类词，misty，boyish 为三类词，rapidly 为四类词；（3）更重要的是，语言的结构体现在语言单位的排列上面，我们有可能根据位置和形态来决定词类。例如，

The vapy koobs dasaked the citar molently
3 1 2 1 4

7.5. 语言形式的描写

对任何一种语言进行描写和分析可以有两种不同的方法。

一种是把言语的较大的单位作为出发点，例如一整段话语；然后把它逐步进行分解成为各个组成部分，如句子和短语；然后再把它们分解成为更小的单位，如具有不同功能的词和词组。这样不断地找出语言的更小的"建筑砌块"，包括各个语音单位。这种方法把句法放到优先的地位，一般用之于研究分析家自己的母语；因为它依赖于分析家对母语的语言能力（linguistic competence），和一个人学习另一种外国语所采用的方法大不相同。

另一种方法从语言的细节开始，学会辨认、发出和写下各种元音和辅音，并在单词和短

语中使用它们。最后才了解单词怎样组成句子，掌握怎样把小单位组成更长的句子的规则。这种方法和一个人学习外语的过程大致相同。

这两种方法，一种"从上而下"，一种"从下而上"，都能有助于我们解决语言问题，但人类语言学家通常使用的是后一种方法，因为他们要研究语言不是母语，而是一种他们知之甚微的外国语。他们只能"从下而上"，一步一步地发现语言的结构，从而提高他们和本地人的交际能力，以了解他们的语言和文化。结构语言学或描写语言学继承了人类语言学的传统，把注意力集中在语言形式的描写。

7.5.1. 语音描写

语言描写有语音、形态和句法三个层面。语音的描写是研究得最充分和最透切的一个层面。这是因为：

1. 人类语言学家在调查处于灭绝边缘的土著语言时必须收集连篇累牍的本族语的资料，这需要用到语音知识，以准确而详尽地进行标音。便于研究发音不同所导致的意义差别。

2. 人类语言的声音系统很不一样，而我们所使用的字母又不足以记录这些声音，如美洲西北岸的一个印第安族的名字是 Tlingit（目前常用的叫法），但不同的文献还称之为 Thlinkit，Tlinkit，Thlinkeet，T'linkets，Klen-e-kate 和 Klenee-kate 的，要把这些声音记录下来必须有一套完善的标音系统。国际音标在 19 世纪后期的制订不但统一了标音系统，而且大大促进了语音的描写和分析。语音学正是在这个基础上诞生的：它所使用的辨别和描写声音的方法使语言学家能够对任何语言的声音进行客观描写，并提出一张详尽的清单。所有人类语言的声音都可以根据相同的基本原则来描写。

3. 从解剖学和生理学的角度对声音发生的研究比较成熟，语音学家的认识也比较一致。

但是语言学家感兴趣的不仅是声音是怎样发生的，他们更感兴趣的是研究一种语言的语音系统或内部组织。这就导致音位学的出现。音位分析牵涉到使用各种方法去发现语音系统，识别音位或声音的功能单位和范畴，弄清音位内部的变体（allophone），找出一种语言所特有的排列、组合和选择的规则。

我们可以从音位分析中看到分布原则的应用。例如：

1. 在相同环境里，两个语音上不同的"声音"可以区别出不同的单词，如英语的 [l] 和 [r] 可以区分 lamb：ram，lot：rot，light：right 等等，我们就把它们称为音位/l/和/r/（方括号为语音符号，斜杠为音位符号）。但在别的语言（如汉语和日语）里，[l] 和 [r] 并不能在相同的环境里区别不同的单词，所以 [l] 和 [r] 的差别并非音位的差别。不能在相同环境里出现的语音单位（因而也就不能区别不同的单词）就是处于互补分布的关系。英语的边音 [l] 有清暗之分，在元音之前念清 [l]，而在辅音前或词尾则念暗 [ɫ]，这可以说是同一个音位在位置上的变体。所以在有些语言里，没有必要把 [l] 和 [r] 看成是两个音位，就等于在英语里，没有必要把清 [l] 和暗 [ɫ] 看成是两个变体一样。但在有些语言（如俄语和某些波兰方言里）它们却是不同的音位。

2. 各个音位之间都有组合关系和聚合关系。英语中的/p/，/b/，/l/在不同的环境里处于聚合关系，如/p/在 pet 里居首位，可为/b/和/l/所替换（试比较 bet 和 let），处于第二个位置的/e/又可为/i/和/ɔ/所替换（试比较 pit 和 pot），处于第三个位置的/t/又可为/k/和/n/所替换（试比较 peck 和 pen）。这样我们就可有下面的一个二维矩阵：

/p/	/e/	/t/
/b/	/i/	/n/
/l/	/ɔ/	/k/

这个矩阵的含义是处于上端的/pet/的任何一个音素都可以为第二和第三行的相同位置的音素所替换，这就形成了一种聚合对立的关系。同时矩阵还告诉我们，替换的结果是七种不同的音素组合：/pet/，/bet/，/let/，/pit/，/pot/，/pen/和/pek/。

我们实际上是把/pet/作为“焦点”来观察聚合关系。但我们也可以不设任何焦点，让第一列的任何音素和第二、第三列的任何音素形成聚合关系，这就有 bin，lick，lock 这些词。但是这样的延伸还会导致一些非英语的单词，如：/bik/和/lɔn/。这些词是从语音上可接受的“词”，但是在现实中并没有“实现”（actualized）。语言中有许多没有实现的组合，使话语产生羡余。①

3. 语言中的每一个音位都有一些特殊的、不同于其他音位的语音特征，可称为区别性特征。这是 Jacobson 和 Halle（1951）首先提出来的，他们（以及布拉格学派的其他成员）所使用的术语不大好懂，主要是建筑在声学的，而不是语音的基础上的。但是他们在研究语言时所采用的理论方法是结构主义的，其目的是发现足以区分任何语言的音位的一些对照特征。他们提出 12 种特征，世界上各种语言的语音系统都可以按这 12 个或其中的一些特征分析。他们认为过去所说的“特征的复杂性”其实“主要是人们的幻觉，”因为这些复杂的现象都可以归结为一些对立。以英语的塞辅音为例，

表 7.1　　英语塞辅音的语音特征

	/k/	/g/	/n/	/p/	/b/	/m/	/t/	/d/	/ŋ/
双唇音				+	+	+			
软颚音	+	+	+						
齿音和齿龈音							+	+	+
浊音	−	·		−	+		−	+	
鼻音	−	−	+	−	−	+	−	−	+

从表中可看出，每一个音位在 5 个变量特征中起码有一个是不同于其他的音位的。在这里还必须进一步指出，这些特征还有功能性与非功能性之别，如/k/和/g/，/p/和/b/，/t/和/d/的语音对立主要表示为浊音的正负值，这个清浊音的对立就是英语塞辅音的最小的功能对立，它是一个区别性特征。另一方面，/n/和/k/，/g/的对立，/m/和/p/，/b/的对立，/ŋ/和/t/，/d/的对立主要表示为鼻音的正值。/n/，/m/和/ŋ/这三个鼻元音虽然也可以发成浊音，但是这在英语语音结构里并不重要，因为在英语里鼻音化本身预设了、决定了浊音的发

① Redundancy，另一种较流行的译法是“冗余”。

生。在聚合对立相同的位置里，英语里没有清鼻音。所以浊音和鼻音一起时，它是非功能性的。这种方法的优点是它能使我们更系统地、更经济地说明某些音位的分布，例如英语里有很多词的开始两个位置是/sp/，/sk/，/st/，但是却没有以/sb/，/sg/和/sd/开始的词。这6种语言事实是有内在联系的，只需要一条规则就能说明："在s ____的环境里清辅音和浊辅音的差别是非功能性的。"

7.5.2. 形态

形态（词法）分析是语言分析的另一个层面。词素（morphemes）是语言中最小的、不能再切分的、有意义的单位。从语言的结构的角度来看，词素分析和音素分析一样，也受到分布原则的支配：

1. 音素有音素变体，词素也有词素变体（allomorphs），例如英语名词的复数词素是/s/，但根据和它组合在一起时词素的语音形式的不同，有时又发成/z/和/iz/。它们的分布受下列规则所支配：(i) 如果复数的结尾是"咝音"，用/iz/；(ii) 如果复数的结尾为浊音（包括元音），用/z/；(iii) 如果复数的结尾为清辅音，用/s/。

2. 词素可以作为一个分布单位来对待。Lyons (1968) 指出，词素的定义里并没有规定词素一定是词中的一个可辨认的切分成分。因此我们可以说英语的 worse（较坏）有两个词素，一个词素和 bad（坏），worst（最坏）共享"坏"的意思，另一个词素和 taller（较高），bigger（较大），nicer（较好）共享"较"的意思。这可以用等值分布的比例关系来描述：

bad : worse : worst = tall : taller : tallest

这个比例关系表示 worse 和 taller 在语法上一样，都是形容词的比较级，因此，

John is { worse / taller } than Michael

但是 worse 和 taller 在意义上是不一样，用它们来修饰名词不能不引起意义上的改变。我们可以用符号的形式来表示这种分布上的比例关系，用不同的字母代表一个不同的词：

A : B : C = D : E : F

如果我们把它分解成因子，就有下列的代数比例：

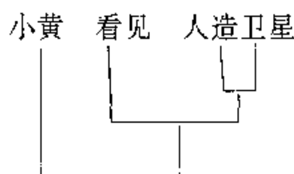
ax : bx : cx = ay : by : cy

这就是说每个词都分解为两个因子。第一个因子是 x 和 y 的区别：方程式左边的词都是 x，右边的词都是 y。第二个因子是 a、b、c 的区别，方程式左边的第一个词相当于方程式右边的第一个词，等等。这些因子也就是词素。

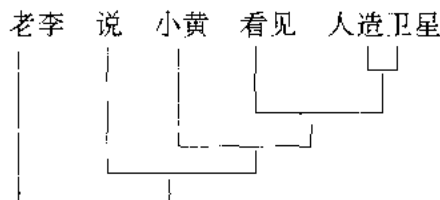
3. 词素有两大类：一类是词根（bases）词素，另一类是附加（affixes）词素。一个语言的词汇就是通过这些词素的组合而产生和发展。语言学家还发现不同语言的词和词素存在着一定的比例关系，在有的语言里，一个词素就是一个词，这就是孤立语，或称分析语。在有的语言里，词根词素和附加词素的关系十分紧密，以致词缀成为词的一部分，如拉丁语 amo（我爱）。amas（你爱）。这就是曲折语。间乎两者之间的是黏着语。Greenberg（1968）指出，词和词素的关系可以表示为一个指数，即词素的数目除以词的数目，其结果就是词的平均词素数。他发现爱斯基摩语的平均数最高，3.72。他还进一步认为，平均数在 1.0 ～2.2 之间的是分析语，2.2—3.0 之间的是综合语（即黏着语）3.0 以上的是多式综合语（即曲折语）。汉语是典型的分析语，英语也属这个范畴；而土耳其语则是典型的多式综合语。在研究儿童语言发展时，我们往往还使用话语平均长度（mean length of utterance，简称 MLU）来作为语言发展的指标。话语平均长度通过计算句子的平均词素数而取得。一般的儿童语言习得的研究者还认为，成人对儿童谈话时使用一种保姆式语言（motherese），其话语总比儿童的话语略为复杂。话语平均长度就是用来衡量这种复杂程度的。当然用词素来衡量语言的类型和语言的复杂性必须十分小心，因为语言学家对什么是“词”并没有统一的说法。有的语言（如英语）把词定义为两个空格之间的字符串，但有的语言（如汉语）就没有那么简单，字和词的界限并不那么明确，切分词的原则也不那么一致。

7.5.3. 直接成分分析法

“直接成分”（immediate constituent，略为 IC）是 Bloomfield（1933）首先提出的，和传统语法中主语和谓语的概念相似，但是它是组合和聚类关系必然的引申：句子由一定次序的有意义的字符串组成，它们包括由词组和词构成的单位，每一个单位都是这个句子的成分。这些单位是连续分割一个句子的组成部分，故称为直接成分。所以一个句子不仅是要素的线形排列，而且是由不同层次的直接成分组成的，每一个低一级的成分都是高一级成分的一部分，故朱德熙（1980）称之为“层次构造”，例如：



这个句子由不同层次构造组成，假如我们要说“老李说小黄看见人造卫星，”那就需要更多一个层次：



直接成分分析法把一个句子分割成许多大大小小的片段，一个大片段又包含一些小片段。直

接成分分析法把人们怎样形成和理解长句的手段加以形式化。一个长句的结构和一个不能再分割的短句的结构基本上是一样的，这些不能再分割的短句可以称为基本句型。长句无非是基本句型的扩展式，例如“老李”可以扩展为“满头白发的老李”，“人造卫星”可以扩展为“我国发射的人造卫星。”

另一种表示直接成分的方法是用括号：

((小黄)(看见(人造(卫星))))

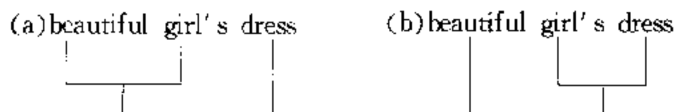
((老李)(说((小黄)(看见(人造(卫星))))))

我们可以在分布的基础上进一步考察直接成分分析法，句法结构可以按照它们和成分的分布分成两大类：向心结构（endocentric construction）和背心结构（exocentric construction）。向心结构是其分布和至少一个成分是相同的结构，也就是说具有分布上等值的关系；所以（a）“满头白发的老李”是向心结构（满头白发的老李＝老李），(b)“广州和北京”也是向心结构（广州和北京＝广州＝北京）。不是向心结构的结构都是背心结构，也就是说它们没有分布上等值的关系。例如汉语的“凉的”（凉的≠凉，凉的≠的），英语的介词短语 in Vancouver (in Vancouver≠Vancouver，in Vancouver≠in）都是背心结构。上述向心结构的（a）和（b）也有所不同，(a）表示的是偏正关系，而（b）表示的是并列结构。成分分析法在汉语语法方面的应用请参考朱德熙（1962）。

直接成分分析法有助于解决句法歧义，例如“我们三人一组”有两个意义：(a）是我们这群人每三个人分为一组；（b）是包括说话人在内的我们三个人分成一组，其他人分成另一组。



又如英语的 beautiful girl's dress 也是歧义的，beautiful 可以修饰（a）girl，也可以修饰（b）dress：

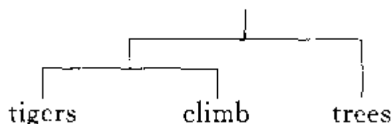


但是直接成分分析法不能解决所有语义的歧义问题(Lyon 称为因素分布上的分类)，例如 They can fish 也有两层意思：一是“他们能钓鱼”(can 为形态助动词，表示“能够”)；二是“他们把鱼造成罐头”(can 为及物动词，表示“做罐头”)。

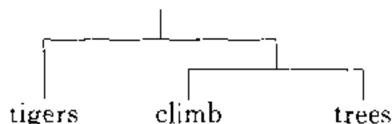
Halliday（1985）认为直接成分分析法在括号里只容许两个成分，并不可取。他认为打括号有两种方法：一种是只要能打括号的地方都打括号，这可以说是最大限度的括号处理(maximal bracketing)，其结果是赋予一个句子以最大限度的结构。这就是直接成分分析法；但是还有另一种方法，在不得不打括号的地方才打括号，这可以说是最低限度的括号处理（minimal bracketing），也可以说是排列成分分析法（ranked constituent analysis)。例如 tigers climb

trees（老虎爬树），用直接成分分析法处理有两种结果（a）和（b）：

(a)

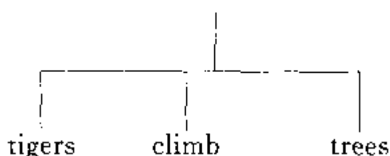


(b)



（a）回答的是 What do tigers climb?（老虎爬什么?），而（b）回答的是 What do tigers do?（老虎干什么?）用排列成分分析法处理的结果是（c）：

(c)



（c）回答的问题是 What have you to tell me?（你有什么要告诉我的?）。由此可见，最低限度的括号处理只处理那些具有某些相对于更大单位而言的功能的成分，它实际上是功能的括号处理。因此研究语言功能就不能采取最大限度的括号处理，而必须采取最低限度的括号处理。我国学者冯志伟（1991）也指出，从自然语言处理的角度看，二分法也有严重缺陷：

1. 二分法不是到处行得通的，特别在汉语中，许多语法形式看来宜于采用多分法，例如双宾语结构"给弟弟一本书"中的"给"有两个宾语，兼语式结构"请他做报告"中的"他"既是"请"的宾语，又是"做报告"的主语，把它们分为三部分来处理，更为清楚。

2. 采用多分法可以在自然语言处理中减少编制程序的工作量：一些长句子，如果采用二分法，层次会多至十层八层，计算机在处理这样多层次的树形结构时，需逐层进行，运算量很大，而采用多分法，大大减少了层次，提高了自然语言计算机处理的工作效率。

3. 采用多分法可以抓住句子的主干，把句子的格局清楚地显示出来，便于检查和研究。

成分分析法是语言形式化处理的不可缺少的手段，Chomsky 所提出的短语结构语法就是在直接成分分析法的基础上发展出来的。

7.6. 型式分析

型式分析（paradigm analysis）[①] 是一种传统语法常用的分析方法。一个型式就是一些语言要素的集合，在集合里，要素的一个部分保持不变，而在另一部分里每个成员都不同。这种分析程序主要是把要研究的要素区分为"同"和"异"。这种方法可以追溯到古希腊亚理士多德派为代表的"整齐论者"（analogists）和以斯多葛派为代表的"参差论者"（anamolists）之争。整齐论者认为语言是有规律性的，可以归结为一些型式，例如下表为拉丁语的两个动词的词法变化表。

① Paradigm 在希腊文里是"模本"、"范例"的意思，但在语言分析里，还指词法变化表。词法变化表就是归纳了各种形态变化，供人模仿的"型式"。

表 7.2　　拉丁语两个动词的词法变化

amo	我爱	laudo	我赞扬
amas	你爱	laudas	你赞扬
amat	他、她、它爱	laudat	他、她、它赞扬
amamus	我们爱	laudamus	我们赞扬
amatis	你们爱	laudatis	你们赞扬
amant	他们爱	laudant	他们赞扬

两个动词的型式都是一样的，第一部分 am- 和 laud-不变，然后第二部分的六个项目均有所不同，但是不同的方式是一样的，这就形成一条简单而全面的规则。

单数		复数	
-o	我	-amus	我们
-is	你	-atis	你们
-at	他、她、它	-ant	他们

其实人类的经验都是用这种方法来分类的，每一种新的经验都按照它和过去的经验的关系而区别为“同”或“异”。这也是我们在语言描写时进行概括的基本的定性方法。我们所做的无非是把一大堆数据（语料）进行归类分析，然后作出表述。

这种分析方法在现代语言学的语言普遍现象和类型学研究中得到广泛使用。在可能的情况下，还附以定量的描写。例如 Greenberg 在研究辅音丛的蕴涵性普遍现象（implicational universals）时，发现词首辅音丛在所有的语言里都蕴涵着词中辅音丛，但有词中辅音丛的却不蕴涵词首辅音丛。这就可以归纳为：

	词中辅音丛	一词中辅音丛
词首辅音丛	+	—
一词首辅音丛	+	—

这就把世界的语言分为四类，而有三类都可找到例证：

1. 既有词首辅音丛又有词中辅音丛的语言（如英语）
2. 有词中辅音丛而无词首辅音丛的语言（如泰米尔语）
3. 既没有词首辅音丛也没有词中辅音丛的语言（如夏威夷语）
4. 有词首辅音丛而无词中辅音丛的语言（无）

另一个常为人引用的是 Greenberg（1963）关于词序的研究。每一种语言都有主语（S）、动词（V）和宾语（O），虽然有些语言的词序比较自由（如拉丁语），但是每一种语言都有其正常的词序，像汉语和英语都是 SVO。这三个成分共有六种组合，但 Greenberg 发现有两种组合（OVS 和 OSV）并不存在，还有一种是很罕见的（VOS）。Utan（1969）所提供的比例数字是

一个很好的补充：

VO（宾语在动词后）的语言		OV（宾语和动词前）的语言	
SVO	35%	SOV	44%
VSO	19%		
VOS	2%		

由此可见，世界上大部分（79%）的语言都是主语在前的。

型式分析法牵涉到三个因素，(a) 维度（dimension），(b) 属性（attribute），(c) 语言在"属性空间"（attribute space）的分布。例如辅音丛的分析有两个维度，每一个维度都有其属性，在这里是有还是没有词首辅音丛，有还是没有词中辅音丛。在类型分析中，每一个维度都是逻辑上独立的，而维度下的属性都有它的值（总是二或多于二），在本例里是有或没有。逻辑上独立指的是：一个属性的任何值可以和别的维度的任何属性的任何值组合起来，构成全部的属性空间。所以每个维度的值的数目相乘就可以得到全部类型的数目，在这里是 4。也可以有一个维度的类型。例如世界上的语言都有元音的音段，所以这是一个单维度类型，其属性为元音性，有两个值：有或没有。世界上的语言都有元音，属于一类；世界上没有语言属于另一类，即没有元音的。

在语义分析中的成分分析法（componential analysis）[①] 其实是把特征分析和类型分析结合起来。例如：

	"男性"	"女性"
"成人"	man	woman
"青年"	boy	girl

这个表表示了两个维度：一个是性别，另一个是成年程度。整个表其实还表示另一个维度——人类，以区别于非人类。因为没有和非人类比较，故没有给出。在几个词的意思可以表示为以下的特征：

man：　＋人类＋成人＋男性
woman：　＋人类＋成人－男性
boy：　＋人类－成人＋男性
girl：　＋人类－成人－男性

这种方法通过比较一些语义对立因素来分析词的意义，是以二元对立为基础的。它进一步发

① 即 4.6 的二元属性分析，其理论上的基本考虑见该节，这里不重复。

展成为语义场的理论，而且被运用到Chomsky的标准理论里，成为选择性限制（selection restriction）原则。但是并非所有的意义都能那么简单归结为二元对立关系的，Slobin（1979）指出，“并非所有意义都能‘分解’为这些特征的，例如把red（红色）和green（绿色），beautiful（美）和ugly（丑），acquaintance（老朋友）和friend（朋友）区别开来特征是什么？……试对chair（椅子）一词作特征分析，就会发现它属于一类边沿模糊、往各个方向发展的词。如果椅背较低，椅脚较长，它就成为stool（凳子），如果座位较宽，它就成为bench（板凳）。”他还特别引用哲学家Wittgenstein关于games（游戏）的论述来说明，一个在不同上下文的词的意义难以用统一的特征来分析，例如board games（棋类游戏），card-games（纸牌游戏）Olympic games（奥运会）。

7.7. 话语分析

话语分析（discourse analysis）亦称语篇分析（text analysis），对这个领域的研究就是语篇语言学（textlinguistics）。一般来说，“话语”偏重于指口头语言，而“语篇”则着重于指书面语言，但是我们在这里把“话语”和“语篇”都用来指口头语言和书面语言，而不再细分，因为讨论的着重点是它们的研究方法。另外，英语的utterance和discourse有时在汉语里都叫做“话语”，其实utterance是discourse中的更小的单位——语句。但是“语句”有时又容易和“句子”（sentence）混淆。所以我们把utterance和discourse都叫“话语”，如果需要区分时，我们才把前者叫做“语句”，后者叫做“语段”。

Harris（1960）在他给《结构语言学方法》所写的序言里对分布法作了两点重要的补充：一是建议用核心句和转换规则来描写语言；二是强调以前的语言分析并没有超出句子的范围，而已知的方法并不容许描述句子之间的结构关系。他提出用分布法的框架来进行话语分析，把出现在相同的语境中的词（或语素）看成是等值的。两个等值的词并不要求它们在意义上是等值的，只要求“出现在相同的环境。”但是这种方法只能用来处理非常简单的句子关系，碰到复杂的句子就不好办，于是他提出转换的方法。

Harris所提出的方法似乎并没有引起大家的注意，不过语言描写不能局限在句子以内，而必须伸延到句子之间的关系，却是很多语言学家的共识。研究话语分析的代表人物之一的Beaugrande（1991）编辑了一本从话语的角度看语言理论的文集：《语言理论》，列举了Saussure，Sapir，Bloomfield，Pike，Hjelmeslev，Chomsky，Firth，Halliday，van Dijk & Kintsch和Hartmann等人的论述，可见不同的学派的语言学家都觉得话语是不能忽略的一个领域。

话语分析可以从定性和定量两个不同的角度来研究：

7.7.1. 话语分析的定性研究

7.7.1.1. 话语分析的理论模型

话语分析的定性研究可以从理论模型和描写框架两个方面来讨论。

1. 早期的研究从Bloomfield和Harris开始，强调的是话语结构的分布。在Chomsky的

转换生成语法影响下，又有人把生成语法的范围延伸到句子以上的语篇结构的研究，例如使用“语言能力”和“语言运用”那样的概念去分析话语。但是有些语言事实是显而易见的，如代词的意义只能从语篇的一些句子的“成分”的关系才能理解，因此 Chomsky 的以句子为中心的看法便受到置疑（见 Rieser，1978）。而从话语分析的角度来看，语言运用的说法也不见得很妥当，Widdowson（1979）指出，“语言运用实际上是一个把语言能力都不能解释的所有的东西堆放在一起的残余范畴，这意味着这些东西都是不完整的、不规则的；而语言能力所能解释的系统的特征则另外放在一个可以称之为说话人语言知识的仓库里。”在后结构主义语言学派中，Pike 的法位学主张对话语结构的影响进行研究，集中在观察这样的一些问题：为什么在某些语境里要用一个长句来代替一段话或要用几个短句来代替一个长句？在什么时候一个意念的结构会编码成几个表层的配置？衔接（cohesion）在话语中起了什么作用？法位学强调在分析表层结构时不要忽略“深层/语义/意念”结构，用型式而不是用规则来表示语言信息，注意考察语言的分层结构，对日后的话语分析理论的发展起了重要的作用。布拉格语言学派对话语分析也作出了贡献，他们认为一个句子应该从三个角度去对待：语义、语法、功能。所以 John explained the problem 这个句子，从语义上看是施事者—动作—目标；从语法上看是主语—动词—谓语；从功能上看是主位—过渡—述位。Firbas（1975）由此提出了“交际能动力”（交际能动力，communicative dynamism）的概念：“交际能动力是交际活动在发展（开展）要传递的信息时所表露的一种特性，它对这种发展起到推进的作用；换句话说，交际能动力这种特性可以表示为对发展话语的贡献程度。”由此可见，交际能动力在不同程度上对不同的句子成分赋予主位或述位的性质，这就显示了句子的不均匀的分布，而且说明对句子的功能分析是话语分析的重要环节。Halliday 的系统功能语法在两个方面对话语分析作出贡献：一是强调文化语境（决定有哪些潜在的行为可供选择）和情境的上下文（决定从行为备选中作出什么真正的选择）；二是提出从“可以做”（can do）到“可以表示”（can mean）和“可以说”（can say）的公式，这就是说，潜在的行为可以通过选择潜在的意义转化为潜在的语言。Austin 和 Searle 的为人熟知的言语行为理论把话语的修辞功能形式化，对话语分析模式的建立也起了很大的影响。所有这些研究都可以说是前话语分析理论研究。

2. 话语分析模型研究是从欧洲语篇语言学的建立开始的。有三种值得注意的倾向：

第一种是企图把生成转换语法引入话语分析，如 Petöfi（1975）提出了语篇结构/世界结构的理论，试图把生成语义学扩展到语篇层次的研究。从方法论的角度看，他的理论主张(a)语篇中的语句是基本的语言单位；(b)分析手段应该包括话语中的语音、句法、逻辑—语义、语用等方面；(c)语言描写的目标是操本族语者对上述几个方面的知识；(d)这些知识可以用明示的规则系统来表示。要达到这样的目标，就必须首先用句法/内涵语义特征来表示自然语言的语篇，然后用世界语义/外延意义特征去解释句法/内涵语义，最后用这两类特征来比较语篇。内涵语义和外延语义的描写都含有语用方面的描写。要把内涵语义转换为外延语义必须使用生成转换语篇语法。

第二种是从心理语言学的角度来研究话语分析，如 van Dijk 的微观/宏观结构理论。其基本出发点是话语处理是输入对所赋予话语结构的函数。句子的形态和句法结构虽然在理解中起了重要作用，但是除了某些语体上的例外，这些结构并非储存在长期记忆里，例如要解释

某个中东政治家的演说的几句开场白当然需要理解各种形态和句法结构,但要有效率地理解,必须首先赋予一个话语结构（开场白)。van Dijk 等（1978）提出这样的一个理论框架：

(i) 话语理论，它包括：

a）话语语法。起码要有

——关于句子和一系列句子的语义表征（命题）理论（微观结构)；

——关于全面的话语结构的语义表征理论（宏观结构)；

——把微观结构和宏观结构联系起来的理论；

b）更为一般的（非语言的）话语结构理论和对不同种类话语的特殊理论；

(ii) 话语结构处理的理论或模型，特别是关于语义信息、记忆储存、记忆转换、检索、产生和使用的处理。

(iii) 更为一般的复杂认知信息处理的理论，把处理话语的能力与接受视觉输入后对复杂的事件和行动的感知和记忆能力联系起来，与在肉体和心灵上计划、组织、执行复杂行动（如推理和解决问题）的能力联系起来。

van Dijk 所说的在微观结构层面上赋予语义表征——“命题”和一般的理解不一样：在把一个句子转化为几个命题时，语义、语用、语体因素，甚至认知和社会因素都在起作用。例如“ He is an ardent reader of the Daily Telegraph”（他是每日电讯报的热情的读者）就可以产生一些命题，如“He is a highly literate intellectual”（他是一个文化程度很高的知识分子)，“He is English-speaking”（他说英语)，“He is Conservative in political affiliation ”（他的政治信仰是保守党)。这已经不是单从语言结构的角度来看命题了。

在 van Dijk 的模型里，话语的全面的意义是由语义宏观结构来表示的。这些结构也用命题来表示，因此我们需要语义映射，即用宏观规则来把微观结构和宏观结构联系起来。这些规则的特征是：它们受话语的一系列的命题（微观结构）所限定，它们的功能是减少或组织信息，也就是说，在某种特定的条件下减少或合并这些命题。宏观规则具有递归性，能够在越来越抽象化的层次上生成更多的宏观规则。对宏观规则的一般的制约是：不能减去那些在话语中成为后来命题的预设的命题。这些规则有四条：

第一条：减少（ Deletion ）

我们可以在一系列命题中减去那些只表示话语指称的附庸性的命题（一般来说，不影响对后面命题的了解)，如：(Mary played with a ball. The ball was blue) ⇒ (Mary play with a ball)。球的颜色对了解后面的命题无甚影响。

第二条：概括（ Generalization ）

我们可以在一系列命题中用一个能够定义微观命题的上义命题来代替这些命题，如：(Mary played with a doll. Mary played with blocks,. . .) ⇒ (Mary played with toys.)

第三条：选择（Selection）

我们可以在一系列命题中略去那些已经为另一个命题所表示的关于事物正常状态、成分、结果的命题。如：〈I went to Paris. So, I went to the station, brought a ticket, took the train...〉⇒〈I went to Paris (by train).〉

第四条：建立（Construction）

如果一系列命题表示的是代替它们的宏观命题的正常状态、成分、结果，我们就可以用一个命题来代替后面接着出现的命题。如〈I went to the station, bought a ticket, ...〉⇒〈I traveled (to Paris) by train.〉

第三种倾向是提出各种综合型的语篇理解模型。例如Schmidt(1978)提出语篇接受模型，认为接受语篇的人必须一开始就了解所看到的语篇属于什么类型，然后才能正确地理解语篇，这个过程牵涉到下面的几个阶段：

1. 语篇的理解从心理/生理感知过程开始，但是它又受个人和社会因素的影响，这些因素决定了理解语篇的复杂的操作过程。

2. 这个感知过程依赖于一个"意义常量"准则：我们期望我们所听的（所读的）东西是有意义的，我们对输入消息的分析是为了和这个准则保持一致。我们分析所采取的途径和类别都是为了达到这个有意义的目标。

3. 接着而来的句法分析过程有两种特殊的形式：一是把赋予语篇的结构切分成更基本的单位；一是从句法上解除歧义，使这些单位组合成句法上连结起来的复杂体。

4. 把句法分析和语义分析结合起来，对句法分析的结构赋予概念结构。这个复杂的过程又牵涉两个环节：(a) 线形的内涵解释，它使解码人能够把输入和记忆中的语法结构、认知结构匹配起来，这就是同语篇（co-text）和语境（context）。(b) 外延解释，它使解码人能够把作出内涵解释的结构和真实的或非真实的世界，即指称系统匹配起来。

5. 从内涵和外延解释过程中建立一个在某一时刻和读者有联系的语义结构或"世界"。

6. 对所解释的语篇结构的这个世界可以修饰和扩大，以达到更高程度的灵活性。

(5) 和 (6) 两点主要是用 van Dijk 所说的宏观结构对信息进行简约和系统化处理。按照他的理解，宏观结构是在理解过程中建立的，而且储存在记忆里，作为提取语义信息的提示。储存在记忆中的宏观结构并非一成不变的，而是结合到所谓"事实与知识系统"里，所以任何信息过程的"理解"都有可能修饰储存在接受人的记忆中的整个概念结构系统。由此可见，Schmidt 的篇章接受模型一方面是 van Dijk 的理解模型的一个补充，另一方面它也吸收了Petöfi关于话语单位的内涵和外延分析的。

Hartmann (1953) 力图把各家的话语分析理论熔于一炉。他继承了德国语言学的传统，并提出了比较类型学（contrastive typology），具有重大的方法论意义。在他看来，比较是为了描写、解释（如印欧语研究）、诠译（如 Humboldt）、分析（如词源学）、概括（如普通语言学）。他提出的比较模型是：

表 7.3　Hartmann 的比较模型

描写：称为 A 的某物和称为 B 的某物是/所做的/所活动的/一样的	
比较（中性的）：	A 和 B 相对应
比较（历史的）：	A 源于/成为/涉及 B
	A 是这样的，因为/如果 B 在这样的
分析和解释：	A 是系统地受制约于 B
	A 经过分析后表示出 B 的功能
评估（形式上的）：	A 的一般的结构要素相当于 B 所需的结构要求
评估（诠译的）：	A 和 B 都可表示为 X 的实现

我国学者胡壮麟（1994）的《语篇的衔接与连贯》对语篇研究的发展也作了很好的介绍和分析，值得一阅。

7.7.1.2. 话语分析的描写框架

话语的描写离不开语言特征的描写，在上面已经谈到，这里无须重复。但是这只是第一步，要对话语进行描写还要进一步考察这些语言特征在不同的情景中的作用。Crystal 等(1969) 认为情景可以分解为各种情景制约维度（dimensions of situational constraints），每一个语言特征的作用必须根据这些维度来进行描写。例如甲特征和一个人来自什么地区有关，可以看成是地域差异维度的特征；乙特征和会话参加者的社会关系有关，可以看成是另一维度的特征——地位。

在进一步讨论这些情景制约之前，我们必须首先考虑一些在语体上中立的，即“共核”的话语特征。这些特征是所有话语共享的；也就是说，它们在情景制约方面的差异是随机的。例如一种语言中的音段语音在各种场合里都必须遵守语法和词汇型式规则。这倒不是说这些特征是超语体的，它们也可作为表示语体的手段，但在使用频率的分布上应有显著的不同。例如一个篇章的所有语调单位都是由降调核构成的，这自然是一种语体特征。除了这些特征以外，其他的特征都在某些方面受到情景制约，Crystal 等人认为，应该有 8 个方面，可分为 3 大类：

A. 个性（Individuality）
　方言（Dialect）
　时间（Time）
B. 话语（Discourse）
　(a)［简单/复杂］媒体（口头，书面）
　(b)［简单/复杂］参与（独白，会话）
C. 使用范围（Province）
　地位（Status）
　情态（Modality）
　独特性（Singularity）

A 类都是一些比较持久的语言特征，具有背景性的特征。文体学家不如普通语言学家和描写语言学家对它们本身的研究那么感兴趣，因为它们不大可能受到情景的影响而发生变异。这些特征不像下面其他的特征那样容易为语言使用者所操纵。从 A 类的几个特征看来，我们可以根据语言和情景之间的关系来进行预测：如果我们已知某些非语言特征，我们就能马上预测出某些语言特征；或是根据语言特征也能预测出其他的非语言特征。例如从生理特点来看，我们可以根据发音特征来预测说话人是儿童还是成人；而从地理特点看来，我们可以根据一个人的方言来预测他是什么地方的人。文体学家觉得单了解这些背景性特征是不够的，必须在这些较明确的特征的基础上再进一步考察别的情景因素对话语的制约。

1. 个性。在不自觉的话语中，有某些特征——相对持久的口语和书面语的特征能够把某一个人和其他使用相同语言的人区别开来，这些具有个人特性的特征一般不会在短期内改变，例如一个人说话的音质和书写的字体，或他经常使用的词语。当然一个人也能为了某种目的有意识地使用另外一种声音或另外一种字体（例如在口技里模仿别人的声音），所以我们在这里强调的是不自觉性。这种个人特性也有别于那些在文体中或别的语境中相对地短暂的，通常是有意识的与个性有关的活动（在“独特性”下讨论）。

2. 方言。在这里指的是广义的方言：地理方言标志了说话人的地理来源的特征，阶级方言标志说话人的社会地位的特征。这些特征一般也比较持久，只有在特定的环境里或受到社会压力时，一个人才会改变其方言型式。不过这些特征的型式并非十分系统化的，只能说是一些倾向（如使用某些词汇或某些元音）。

3. 时间。在话语描写中的时间表示历时信息的特征。不管是把语言作为一个整体，还是把语言作为某些个人的语言习惯，这些信息对研究一种语言的历史都是十分重要的。语言的时间特征也是相当持久的。

B 类和话语有关，它的变异情况可以从媒体（口语和书面语），或从说话人的参与情况（独白和会话）的不同角度来研究。B 类的特征和上面讨论的特征不同，可以参照别的特征（如下面要谈到的情态）来考察其变异情况。换句话说，我们必须参照语言在使用时的基本特征来区别其差异：我们所关心的不是这些特征所提供的描写的信息，而是它们所提供的解释价值，对这些语言差异的考察更能说明某些变体的特征。例如在一篇书面语的语篇里出现一些通常只能在非正式的口语中出现的特征，或是在一篇口语材料里出现某些只见于书面语的结构，或是某人把一些会话的特征引入独白里去。

1. 书面语和口语是两种不同的，但又有所重叠的语言和非语言特征。无论从语体的角度，还是从方法论的角度，都值得加以研究。它们的差异首先是非语言的，与交际所采取的方式和材料有关——是在空气中传递的符号还是写在表层上的符号（当然还有其他的方式，如图画、音乐等等）。这个区别对文体学家和普通语言学家都是同等重要的：口语需要在语音/音素层面上进行，而书面语需要在字符/字素层面上进行。两者牵涉到不同的描写框架。从情景的角度看，两者不但有重要的功能上的差别（口语较短暂，而书面语则较持久；口语表示某种个人的接触，而书面语则不同），而且两种变体在形式上缺乏完全的对应。我们很难用传统的书写的方法去写口语，以反映口语的各种差异（例如在书面语中只好省去各种超音段的信息、有些猥亵的话也不好写下来）。有些书面语要说出来，也只能破坏原来语篇的连贯性（例

如要朗读一些没有标点符号的法律文件，就只好把它切分为一些在语篇中不存在的单位）。下表显示了口语和书面语的一些差异：

表 7.4　口语和书面语的差异

口语	书面语
1. 使用听说信道（语音表征）	使用视觉信道（图形表征）
2. 短暂性	半持久性
3. 接受的单向性	接受的全方位性
4. 韵律性（音高、响度、音速、节奏、停顿）和伴随语言手段（面部表情、手势、姿势）	图形手段（图表、标点符号、大写、图画、斜体、标记、画线）
5. 对话人在场（即时反馈）	对话人不在场（通过词语与读者接触）
6. 较多的重复和羡余	较少的重复和羡余
7. 多数为非正式	多数为正式
8. 隐含性（与情景联系、歧义、句子不完整）	明示性（在阅读中建立语境）
9. 题材的随机性，缺乏计划性	完整性、连贯性、简练性、组织性

2. 从参与话语的角度看，有独白（不期望别人作出回应的话语）和会话（参加者轮流说话的话语）。有几个原因使我们把两者的差异和媒介的差异等量齐观：首先是参与和媒介一样，是在较为一般的抽象的层面上实行的；其次是参与和媒介有明确的相互关系：有口语体和书面语体的独白，也有口语体的会话（十分明显的）和书面语体的会话（如填表、交换书信和一些集体游戏）。第三是参与和媒介在功能上有很多相同之处，两者在使用上都有制约：我们不可能对听不到我们说话的人说话；在别人在场的情况下，我们不会说独白；一般我们不会对同一房间的人写信。两者都可以有"移位"或"解释"的作用：例如把媒介作为一种达到目的的手段，而不是目的本身：口语本来是说出来给人听的，但我们可以说出来给人写（如听写）；书面语本来是写下来给人看的，但我们可以写下来给人说（如新闻广播）。这种现象可以称为复合媒介（complex medium）。同样的，独白/会话的区别往往也可以"移位"，在独白中引进一段会话（如说书），或在会话中引进一段独白（如说一个笑话）。

C 类包括使用范围、地位、情态和独特性。这些特征一般都是比较短暂的，可以进行转换的。

1. 使用范围。我们所描写的语言特征可用来识别一段话语的变量是否和某一超语言的语境（如职业性的或专业性的活动）有联系。使用范围的特征对说话人的社会地位和相互关系不一定能提供什么信息；但是不管参加者是谁，这些特征和他们所从事的任务的性质有关，会反复出现。例如语言使用者的职业身份（如律师、医生）对他们所说的和所写的会形成某些制约（或从正面说，提供一套供他们自由运用的语言形式）。使用范围和后面还要讨论的地位和情态一起成为任何语言变体的语体基础。我们通常把这些变体称为某某语言，如广告语言、会计语言、计算机语言、法律语言……，表示"用于某一范围的一套独特的语言特征"。根据其语言特征的性质的不同，这个情景变量可以有不同程度的概括，我们可以说"广告语言"，也可以说"电视广告语言"，也可以说"洗衣粉电视广告语言"。

关于使用范围的概念，还有三点值得注意的：首先，使用范围特征和话语的题材不能混

为一谈。就词汇使用而言，题材只是范围的一个因素，只有在非常专门的情景里，才有预测力。其次，我们一般用“使用范围”的说法来说明会话，但是会话和其他的使用范围不同，传统的职业界限对它无大影响。从语言上看，会话可以出现在或跨越任何有限制的语言用途（如律师和医生在谈家常时的使用范围是一样的）。第三，使用范围的特征和其他的特征（如地位）相比，并非要优先考虑的，只不过在描写话语时，比较容易从使用范围入手。

2. 地位。地位特征反映了和说话人（不管他们来自什么地区）所处的社会地位变化相适应的系统的语言变化。地位特征的变化独立于使用范围特征的变化。“地位”下面的语义场是相当复杂的，它包括了人们在社会量表里的各个位置上进行接触所需要的方方面面的因素，如：正式、非正式、尊敬、有礼貌、服从、亲密、亲属关系、商业关系、一般的上下关系。在地位特征中可以明显地区别出一些领域，各种正式的和非正式的语言是最惹人注目的。Joos提出五种程度不同的语体（僵化的、正式的、商量式的、随便的、亲密的），但是Crystal等人认为还不够成熟，因为尽管我们能够找到完全放进这5个范畴的例子，不能放进的例子要多得多。

3. 情态。情态在语体研究中并没有加以系统地区别开来。情态指的是那些按照话语的特定目标而进行修正的语言特征。话语的目标不同往往会令语言使用者专门采取某一种或某一套特征，最后产生一种可以贴上某种描写性的标签的完整的、常规的口语或书面语的格式。情态独立于使用范围和地位，因为不管语言使用者的职业身份如何，他们的相互关系如何，我们都可以自由选择某种情态。例如在会话的范围内，如果我们采取书面语的形式（可称为“通讯”），那么视方式的不同（如信件、明信片、记录、电报、备忘录）而有不同的语言情态的差别；在科技语言的范围内，如果我们要讨论一个题目，用什么方式（如讲稿、报告、论文、专题文章、教科书）来写作，也会有不同的语言情态的差别。我们通常所说的文学类型(genre)也可以从情态的角度来观察。从局部看，情态明显地是一个形式对题材的适宜性的问题，但有时不能完全这么看，因为一些常规的语言格式是有继承性，例如很多法律文件中的信件的形式。不过鉴于有时使用范围和情态的界限不太容易区分，所以还是有必须强调情态的独立性。例如我们把“体育实况广播”看成是一个使用范围，实际上是把两个理论上的变量混为一谈了。从实况广播的角度看，这是一个使用范围；但是广播中的评述却是一种情态特征（区别于报纸报道、电台综述等）。情态特征可以跨越使用范围，例如我们可以有体育的实况广播，也可以有烹调的实况广播，甚至发射导弹的实况广播。

4. 独特性。除了按照上述几个方面来描写篇章外，仍然可能有一些语言特征并非系统地存在于言语社区或群体当中，而仅见于某些个人。一个人可能在他的话语中显示某些偶然的、个人独特的语言特征，使某些传统的变体产生特殊的效应，例如一个作家在他的诗歌里使用了某些独创的语言。独特性和个性不同：前者是偶而为之的、比较短暂的，可操纵的，通常是有意使用以创造特殊的语言效应的；而后者是比较持久的、连续性的，不易操纵的特征。

根据上面的讨论，Crystal等人认为可以把话语的描写框架归结为对13个子题目的回答。也就是说，除了要传递的消息外，话语还传递了一些什么信息？

1. 它有没有告诉我们什么样的人在使用话语？（个性）
2. 它有没有告诉我们他来自什么地方？（地区方言）

3. 它有没有告诉我们他属于什么社会阶层？（阶级方言）

4. 它有没有告诉我们他在语言的哪个发展阶段使用话语？他有多大年纪？（时间）

5. 它有没有告诉我们他是在说话还是在写作？（话语媒介）

6. 它有没有告诉我们他的说话或写作本身就是目的，还是达到目的的一种手段？（简单的，还是复杂的话语媒介）

7. 它有没有告诉我们话语中只有一个参加者，还是多于一个参加者？（话语参与）

8. 它有没有告诉我们所进行的独白和会话是独立的，还是作为一篇更大的话语的一部分？（简单的，还是复杂的话语参与）

9. 它有没有告诉我们语言使用者是在从事什么特定的职业性的活动？（使用范围）

10. 它有没有告诉我们语言使用者和他的交谈者是一种什么社会关系？（地位）

11. 它有没有告诉我们他在传递消息时心中有什么目的？（情态）

12. 它有没有告诉我们语言使用者是在有意地表示一些个人的倾向？（独特性）

13. 它是否没有告诉我们上面的任何信息？（共核）

7.7.2. 话语分析的定量研究

在某个意义上说，话语的描写和分析所牵涉到的语言特征，都有一个量的关系。假定说，我们在英语的被动语态和主动语态之间作出选择，宁愿说 Jane Austen wrote *Persuasion*，而不说 *Persuasion* was written by Jane Austin，这不存在什么语体（风格）的问题。但是如果在一段语篇里，被动语态反复出现，或者说被动语态的频率大大高于主动语态的频率，我们就有理由说这段语篇具有倾向于使用被动语态的特征。所以 Leech & Short (1981) 说，"看来，文体学家成了统计学家。" Enkvist 等 (1978) 认为，语篇语言学和文体学在广义上所覆盖的范围差不多；但从狭义上看，语篇语言学是研究句子相互关系的文体标记 (style markers) 的。文体标记指的是"任何语言特征，它在语篇内的密度显著地不同于它在语境相关的常模里的密度。所以如果某一特征只出现在某一语篇里，而不见于常模里，它就是一个文体标记；如果某一特征出现在常模里，而不见于在某一语篇里，它也是一个文体标记。如果某一相同的特征既出现在某一语篇里，又出现在常模里，只要它在语体内的密度和常模里的密度显著地不同，它也是一个文体标记。这就是为什么语言文体学经常变为一门定量科学的原因。"

7.7.2.1. 统计学方法

把统计学方法用来研究语言是统计语言学 (statistical linguistics)，用来研究语言文体的是文体统计学 (stylo statistics)，它是在 50 年代问世的信息论的影响下出现的，而且很快就和文学评论中的新批评主义和具有悠久历史的版本学研究结合起来。当时的计算机使用还未十分普及，但这种研究方向已为日后的计算语言学奠定了基础。在这里，我们限于篇幅，不能详细介绍各种具体的统计学方法，而只能涉及一些基本的概念，读者如感兴趣，可参阅 Herden (1960, 1964)，Yule (1968)，桂诗春 (1991)，冯志伟 (1991)。

语言项目在篇章中的出现频率是语言学家都承认的一个重要的特征，Herden 指出，"言语社区成员不但在使用语音系统、词汇和语法系统方面，而且在使用特定的音素、词项和语法

形式与结构的频率方面都十分相似。换句话说，相似的不仅是使用了什么，而且是使用了多少。”频率必须根据一个总体来统计的，具体地说，就是类型和标形的比率（type / token ratio）关系，所以 Herden 把他的数学语言学的教科书称之为“类型—标形的数学”。最基本的方法是按照每个词项的出现频率来统计词汇的频率分布，并用这个方法来研究文体特征。这个方法在“作者考证”研究中被普遍使用。我们可以从两个方面来标志文体的定量特征：

1. 作为比较研究的基础，我们可以从词汇和出现频率（用词的数目来表示篇章长度）的关系，也可以单从词汇或单从出现频率的不同的角度，来表示其特征。

例如从 40 年代开始就有人考察不同作者或不同语篇的类型和标形的关系，标形是一个篇章的全部的单词数，而类型是全部单词中的不同的单词数，例如在统计新闻英语时，一篇材料有 44 000 个标形，6 000 个类型，那么类型/标形比率就是 0.136。这种比率的问题是：样本越大（即标形越多），比率就越小，因为篇章越长，重复使用的单词的可能性就越大。为了寻找一个独立于篇章长度的文体特征标志，Herdan 建议使用类型/标形的对数比率，即先把类型数和标形数作对数转换，然后再求其比率。他使用了这个方法来研究希腊原文新约中的使徒书中三封有争议的书信是否出自使徒保罗的手笔，发现保罗的其他书信的平均类型/标形对数比率为 0.8113，而这三封书信的比率分别为 0.8540，0.8605，0.8794，因而断定非保罗所为。他用同样方法研究新约中圣约翰的书信和启示录，发现两者的比率很接近，和整本新约的比率不同，因而证实了约翰使徒书和启示录出自同一人的假设。Herdan 认为这种简单的方法虽然受到一些语言学家的怀疑，但却有两点价值：

(a) 它可以用定量的方法来代替模糊的趋势的概念，而且说明文字的相对增长率和有机体的增长率是一致的。对数比率不但可以标志一个篇章的各个部分的特征，而且也可以标志同质性的各个篇章的特征；这也就是说，只要“环境”的条件一样，类型/标形的对数比率就是一个常数。

(b) 它在比较文体的相同性或差异性方面有实用的价值。

Yule 所提出的 K 特征（Characteristic K）也是出于同一目的：企图寻找一个不受样本数影响的、表示类型和标形关系的统计量。但是对这个统计量却有不同的解释，Yule 自己认为这是一个词汇分布的统计常量，说明词汇的集中程度。但 Williams (1946) 把它看成是一个表示差异程度的统计量，Good (1953) 认为它表示的是词的重复率，即在一个篇章中随机选择的词会不会是同一本字典里的词的概率。Herdan 认为它不仅是一个文体统计的参数，而且还可以用来标志语言发展某个阶段的词汇和它出现率之间的关系，即概括“语言和言语”的关系。

Yule 所给的 K 特征的公式是

$$K = \frac{S_2 - N}{N^2} \qquad (7.1)$$

$$N = \sum_{1}^{s} r n_r$$

$$S_2 = \sum_{1}^{s} r^2 n_r$$

在公式里，r 与 n_r 分别为词出现率和词汇的频率。Yule 还主张把 K 乘以 1000，以避免过多的

小数。让我们看一个具体的应用 K 特征来研究文体的例子：

英国散文家 Macaulay 于 1825～1842 年间在《爱丁堡评论》上发表了一连串的论说文，纵论时文，Yule 从中选了他讨论 Milton（1825）、Hampden（1831）、Frederic the Great（1842），Bacon（1837）等几篇文章，其代号分别为 A，B，C，D，进行统计。D 采用了交叉选样，然后取其平均数的办法，其他文章则整篇计算：

表 7.5　Yule 对 Macaulay 几篇文章的词频统计

出现率 (r)	词汇频率 (n_r)			出现率 (r)	词汇频率 (n_r)		
	A	B	C		A	B	C
1	851	721	865	25	0	1	0
2	305	233	260	26	1	1	1
3	128	117	150	27	1	1	1
4	73	77	75	28	0	1	1
5	38	34	50	29	0	3	1
6	36	25	35	30	2	0	0
7	20	19	19	31	1	2	0
8	13	20	19	33	0	0	2
9	10	12	16	34	0	0	1
10	6	15	13	36	0	1	0
11	9	5	4	37	1	0	0
12	12	7	6	39	0	2	1
13	4	6	3	44	0	0	1
14	1	4	4	47	0	1	0
15	5	3	6	56	0	1	0
16	3	3	3	63	0	0	1
17	5	1	1	64	0	1	0
18	3	4	2	68	1	0	0
19	3	2	0	73	0	1	0
20	2	1	2	83	0	1	0
21	0	2	0	87	0	0	1
22	1	2	1	93	0	1	0
23	1	0	0				
24	1	2	0	总计	1573	1333	1545

按照所列公式 $N=\sum rn_r$，以 A 为例，$N = 1\times851+2\times305+3\times128\cdots\cdots+93\times0=3923$；$S_2=\sum r^2n_r$，故 $S^2=1^2\times851+2^2\times305+3^2\times128\cdots\cdots+93^2\times0=31489$；而 K 乘以 10000 后＝17. 91。

以下是统计的结果：

表 7.6　　表 7.5 的统计结果

1	2	3	4	5	6	7	8	9	10
文章	单词数	N	S_2	平均数	1000 × 单词数/N	方差	均方差	出现一次的单词（%）	K 平均数
A	1，537	3，923	31，489	2.55	392	13.93	3.74	55.4	17.91
B	1，333	4，149	62，839	3.11	321	27.45	6.12	54.1	34.09
C	1，545	4，061	40.045	2.63	380	19.01	4.36	56.0	21.82
D	1，413	4，022	48，274	2.85	351	26.04	5.10	54.4	27.33

Yule 认为 A 和 B 的 K 值代表了 4 篇文章的两个极端，A 的 K 值最低，而使用的词最多（见第五栏的平均数、平均数越小表示重复率越小）。原来这篇文章是 Macaulay 年青时候，刚刚大学毕业后写的，他自己在文集前的序言也承认，"论 Milton 的文章是作者刚离开大学时写的，其中没有哪一段话是作者成熟判断力所赞同的，它充满了华而不实的、庸俗的装饰。"如果把这 4 篇文章和其他作家的文章比较，则 K 值还是比较低的：

Macaulay A	17.9
B	34.1
C	21.8
D	27.3
Gerson	35.9
Thomas à Kempis	59.7
Imitatio	84.2
St. John：Basic	141.5
St. John：Basic	141.5
St. John：A. V.，all	161.5
St，John：A. V.，selected	177.9

Yule 再进一步考察 A，B，C 中使用频率超过 40 次的名词，就发现它们有很大的不同：

表 7.7　　几个不同文本用词的频率的差异

A. Milton		B. Hampden		C. Frederic	
频率	词	频率	词	频率	词
20	feeling：king	20	country	20	country：troop
22	power	21	act：nation	22	life
23	liberty	22	court：opposition	26	battle
24	people	24	part：power	27	day
26	work	25	government	28	power
27	time	26	day	29	part
30	mind：poetry	27	liberty	33	prince：year
31	character	28	place	34	time
37	poet	29	law：person：war	39	war

A. Milton		B. Hampden		C. Frederic	
68	man	31	army: party	44	army
		36	year	63	man
		39	member: people	87	king
		47	commons		
		56	time		
		64	house		
		73	man		
		83	parliament		
		93	king		

在讨论 Milton 的文章里，真正的高频词只有 man (68)，然后是 poet (37)、character (31) 和 poetry (30)，都是和文学讨论有关；政治性词汇如 king，power，liberty，people，居于单子的最末。与此相反，政治性词汇居于 B 的首位，而且频率甚高，所以重复率远远高于 A。Yule 用同样的方法还了研究 Bunyan 的文体，并从各个方面进行比较。

2. 从词汇在一个整体的各个部分的分布或从某些词项出现的概率来表示其特征。例如 Mosteller 和 Wallace (1964, 1985) 在考证 12 篇匿名为"联邦主义者"究竟是出自政治家 Hamilton 还是出自美国第四任总统 Madison 的手笔时。就采取这个方法。这两位作者都是政论家，而且有很多相似之处，例如统计他们所写的没有争议的文章显示其句子的平均词长分别为 34.5 和 34.6。但两人在使用 while 还是使用 whilst 方面有所不同：在 14 篇公认是 Madison 所写的联邦主义者的文章里，while 一次都没有出现过，whilst 却出现过 8 次；而在 Hamilton 所写的 48 篇文章中，while 出现过 15 次，whilst 却一次都没有出现过。Mosteller 等便决定使用一些标记词来加以比较：例如在名词方面，他们选了 commonly、innovation、war；在虚词方面他们选了 by、from、upon 等。以 upon 为例：

表 7.8　　upon 在 Hamilton、Madison 和其他有争议文章的出现率

在 1000 词中的出现率	Hamilton	Madison	有争议的文章
0		41	11
0—0.4		2	
0.4—0.8		4	
0.8—1.2	2	1	1
1.2—1.6	3	2	
1.6—2.0	6		
2.0—3.0	11		
3.0—4.0	11		
4.0—5.0	10		
5.0—6.0	3		
6.0—7.0	1		
7.0—8.0	1		
总计	48	50	12

两人在使用 upon 的趋势是很不相同：在调查的 Madison 所写的 50 篇文章里，有 41 篇是一个 upon 都没有用的，其余使用最多的两篇也仅在 1.2-1.6%之间；而 Hamilton 则不同，upon 在 48 篇文章里都有使用，它在 32 篇中的出现率在 2.0-5.0%之间。而 upon 在有争议的 12 篇文章里的使用的情况和 Madison 的相同。因此 Mosteller 的结论是除了一篇编号为 55 的文章外，其余 11 篇均为 Madison 所写，可能性为 80 比 1，即 98.8%。

这种方法在计算机的支持下得到很大的发展，如前苏联的学者 Kjetsa 对肖洛霍夫《静静的顿河》的研究确认肖洛霍夫是真正的作者，后来又发现了小说的两篇原稿，经鉴定属肖洛霍夫手笔。(见冯志伟，1991)

英国的 Smith (1983) 不但使用了计算机，而且还使用更为复杂的统计方法来研究文体。他提出了和文学批评中的形式主义和结构主义(新批评主义)相一致的计算机批评主义(computer criticism)，而且认为计算机批评主义不同之处在于要求在更多层次上建立形式化的规则，这些层次是根据各种特定的批评要求而任意制订的，因而能观察范围更广的篇章特征。从形式上看，一篇作品不但有范畴，而且这些范畴还有序列，即有历时的关系，这些关系可以采取一些先进的统计手段，如因素分析中的主要成分分析法，时间序列法(傅立叶分析法)来进行分析。他对 Joyce 的《一个作为年青人的艺术家画像》的第一章进行了主题分析，认为这个主题是围绕火与水两个概念展开的。他把与火和水有关的词找出来：如 burn、burned、burning、fire、heat、hot 和 water、damp、watery、wet 等，然后把整章分为以 500 个词为单位的小段，再分别统计这两组词的分布，就能看出主题的开展：

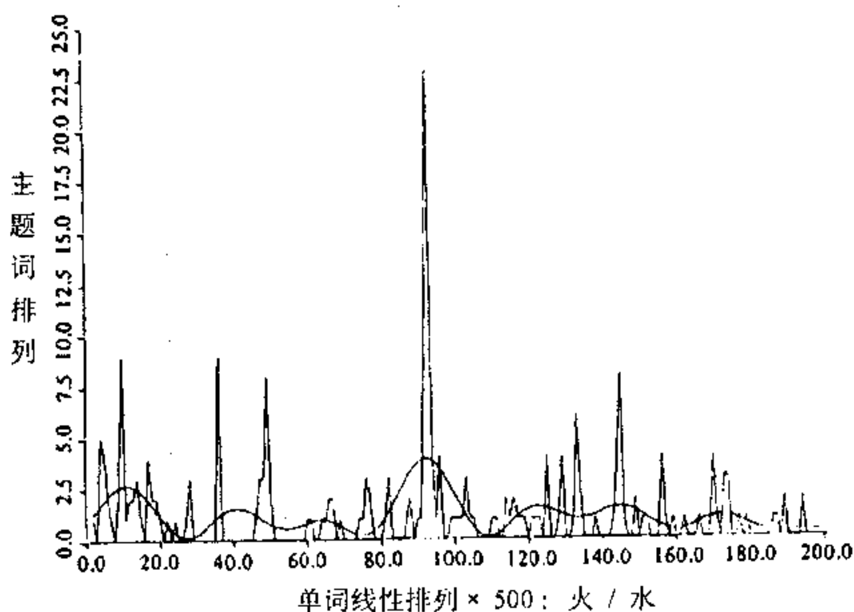


图 7.3 画像第一章主题的展开

7.7.2.2. 语料库方法

60 年代以来，随着计算机的普及，出现了很多英语语篇的机读语料库 (machine-readable corpora)，使话语的分析和研究跃进了一大步。最早建立的是 Brown 语料库 (1961)，称为《布朗大学现代美国英语标准语料库》，当时的语言学潮流是反对使用语料库作为语言研究的

数据库的，但是以磁带形式发行的这个语料库在几年内就卖出了160多套，购置的单位发现语料库不但对语言学，而且对教育学、心理学、哲学、文体学和技术研究都很有用处。1970年英国Lancaster大学开始建立一个与布朗大学语料库相平行的英国英语语料库，并在挪威Oslo大学和Bergen的挪威人文科学计算机中心的帮助下，终于1978年建成，称为LOB语料库。和布朗语料库一样，LOB语料库也是用磁带的形式发行。随着计算机的普及，各种语言的语料库日益增多，而且收录的语料以惊人的速度增长，现在语料库已进入互联网络(Internet)，上网的用户可随时调用。规模最大的是英国Birmingham大学和Collins等出版社联合建立的英语库(Bank of English)，已有两亿一千多万词，每月以500万词的速度增长，而且收录的大部分为90年代以后的资料，包括口语资料（由专人把录音资料整理成文字，每月增长速度为50万词）。英国牛津大学出版社牵头建立的英国国家语料库（BNC，British National Corpus）也收录了一亿词（90%为书面资料，10%为口语资料）。英国的牛津大学计算机服务中心建立的篇章档案库（Oxford Text Archive）收集了超过十亿比特，1500部电子语篇，包括了一些主要作家的希腊语、拉丁语、英语和其他语言的作品，基本上是无偿地供公众使用。正在建立的还有国际英语语料库（ICE，International Corpus oF English)，在20多个以英语为第一或第二语言的国家和地区建立子语料库（每个100万个词）。专用的语料库也纷纷建立，如学生英语国际语料库（ICLE，International Corpus of Learner English)，准备收集9种不同语言背景的英语学生的语料，约100万个词。Quirk的英语用法调查语料库(SEU，Survey of English Usage Corpus)，有100万个词的语料。泰晤时报语料库（TST，Times and Sunday Times Corpus）有3300万词，印度英语语料库（Kolhapur Corpus of Indian English)，有100万词，Leuven戏剧语料库收录了100万英国戏剧语篇资料，IBM在美国的Watson研究中心使用了6000万的语料来研究言语识别。上海交通大学建立了100万词的科技英语语料库，香港科技大学和广州外国语学院联合建立了100万词的计算机英语语料库。在汉语方面，我国在20年代就有陈鹤琴开始进行汉字的词频调查；1974年的“748工程”从两千多万个字中统计出不同的汉字6347个，编成《汉字频度表》；1986年北京语言学院根据200万字的统计，编成《现代汉语频率词典》；1985年北京航空学院和国家语委使用计算机从一千多万字的语料编制了《现代汉语常用字表》（陈原1989，冯志伟1991）。不过这些调查虽然抽样面广，有的也使用了计算机，主要是为了统计词频，而不是为了建立供研究汉语文体用的语料库，这是比较可惜的。

1. 采样。建立语料库首先碰到的问题是采样，最基本的方法是分层随机抽样，即首先按照语料分布确定抽样的范围，然后再在从每个范围里随机抽取样本。样本越小，就越要讲究抽样的代表性。以较早建立的Brown和LOB两个语料库为例，为了更好地比较美国英语和英国英语，这两个语料库的格式是一致的，各有100万个词的语料，从15个领域抽选了500篇语料，每篇有2000个标形。它们的分布如（表7.9）：

新建的英语库采样更为广泛，包括美国英语和美国英语的各种文体，各种小说与非小说的书籍，报章和杂志，个人信件，广告和宣传品等等。光口语资料就有300万词。因为样本较大，也就没有那么讲究随机性。

2. 数据输入。基本上有二个途径：一个是电子资料，计算机在90年代以后的大普及使很多作家都在计算机上使用文字编辑器来写作，写下来的就是现成的电子资料，可直接进入语

料库。牛津大学出版社在英语库建立中负责提供书面资料，实际上都是作家的电子原稿；第二个是使用扫描仪，一些较先进的光学字母阅读（optical character reader）软件准确率越来越高（一般都超过 90%），而且还有一些防错和纠错的功能，大大提高了资料进库的速度；第三个是键盘输入，主要是一些口语资料，需要有专人把它变成文字，然后在键盘上输入。

3. 数据的整理和分析。语料库基本上是进行词频排列，可以以词为单位排列词频，也可以以词频为单位排列词，还可以分不同的领域来分别统计和比较。但是任何一个语料库都体现为类型和标形的关系，究竟它们是一个什么关系？也就是说，词频分布有些什么规律可寻？这是我们首先要探讨的问题。

表 7.9　　Brown 与 LOB 两个语料库的抽样范围和数目

篇章范畴（文体类型）	篇章数目	
	Brown	LOB
A　新闻：报道	44	44
B　新闻：社论	27	27
C　新闻：评论	17	17
D　宗教	17	17
E　技能和爱好	36	38
F　民间传说	48	44
G　纯文学，传记，回忆录等	75	77
H　杂项（主要是政府文件）	30	30
J　学术性（包括科学技术）	80	80
K　一般性小说	29	29
L　神秘性小说和侦探小说	24	24
M　科学小说	6	6
N　冒险和西部小说	29	29
P　浪漫和爱情故事	29	29
R　幽默	9	9
总计	500	500

(a) 较早考察这种关系的是 Zift（1935），他认为，如按频率递减的顺序排列，越排在前面的词频率越大。如果把频率和顺序转化为对数作图，则它们的关系是一条从左上角到右下角的对角直线，用数学公式表示就是：

$$P_r = 1/(10r) \qquad (7.2)$$

r 是顺序，P_r 是顺序的相对频率，所以排在第一位为 0.10，第二位为 0.05，第三位为 0.033，……等等。rP_r 是一个常数：0.1。但是很多实际统计的结果显示词的频率和它的排列位置不完全是直线的，特别是开始的几个词和后面的一些词的递减程度没有那么大。因此不少人对 Zift 定律提出异议，Mendelbrot（1965）主张把这条线修正为开始时略平，然后逐步往下垂的

曲线。Herdan（1960）提出了对数正态模型（lognormal model），也就是说，如果用对数来表示频率，那么一个语料库中的词汇的分布是正态的。Carroll（1971）用对数正态模型观察了 Brown 和 AHI（American Heritage Intermediate Corpus）两个语料库，认为它比 Zift 定律要好。

我们可以先看 Herdan 所提供的一个简单的例子：他根据 Bell 电话公司的 8 万个词的电话会话，统计出在 76，054 个标形中有 738 个词项，然后再统计这些词汇的字母数和音素数。在下表中，pi 为单词所有的 i 单位的百分比，而 xi 为单词的平均频率。

表 7.10　　738 个词项按字母和音素数的频数表

	字母				音素			
每个词所有的 i 单位	词汇		出现率		词汇		出现率	
	pi	$\sum$pi	piXi	$\sum$piXi	pi	$\sum$pi	pixi	$\sum$pixi
1	0.3	0.3	8.0	8.0	0.4	0.4	8.0	8.0
2	3.2	3.5	20.9	28.9	8.1	8.5	36.8	44.8
3	9.8	13.3	25.2	54.1	30.8	39.23	35.0	79.8
4	26.5	39.8	26.2	80.3	23.4	62.7	11.4	91.2
5	19.1	58.9	9.45	89.8	14.6	77.3	4.4	95.7
6	15.6	74.5	4.36	94.1	8.0	85.3	1.6	97.2
7	10.0	84.5	2.5	96.6	6.0	91.3	1.3	98.5
8	6.1	90.6	1.87	98.5	4.5	95.8	0.8	99.3
9	4.2	94.8	0.8	99.3	2.7	98.4	0.5	99.8
10	2.6	97.4	0.46	99.8	1.1	99.5	0.08	99.96
11	1.9	99.3	0.25	99.95	0.14	99.7	0.01	99.98
12	0.4	99.7	0.03	99.99	0.27	100	0.02	100
13	0.2	100	0.01	100	—	—	—	—
$\log_{10}i$ 的平均	0.703		0.494		0.608		0.414	
$\log_{10}i$ 的均方差	0.168		0.218		0.189		0.187	

如果把这些分布作在横座标为对数量表、纵座标为的高斯积分的概率纸上，就可以看到几条基本上是平行的直线，说明分布是对数正态的，如（图 7.4）。

Herdan 由此归纳出一条公式：

$$\mu_j = \mu + 2.3026 j\sigma^2 \qquad (7.3)$$

j 表示动差（moment），标形的分布是类型分布的 1 级动差分布．所以它和类型分布的距离就表示为 σ^2，例如他观察了新约全书和 Macaulay 讨论 Bacon 的文章，其结果如（表 7.11）。

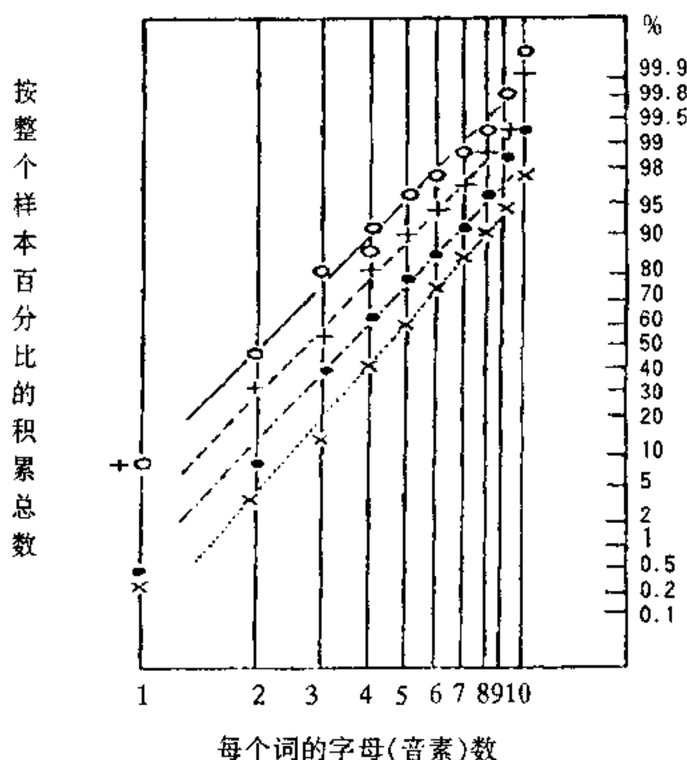


图 7.4　类型分布与标形分布

表 7.11　　新约全书和 Macaulay 的文章的类型和标形统计

	对数平均（以10为底边）	标形和类型之差	对数均方差（以10为底边）	平均均方差	$2.3026\sigma^2$
新约全书					
类型	0.2553	1.7447	0.8587	0.8800	1.783
标形	2.0000		0.9031		
评论 Bacon 文章					
类型	0.1461	0.7290	0.5863	0.5910	0.80
标形	0.8751		0.5963		

由此可见，类型的对数平均加上 $2.3026j\sigma^2$ 便是标形的对数平均，故 μ_j 又是标形的理论分布。应该指出的是 j 不是永远都是 1 级的，以上述的字母和音素统计为例，调查的是词的长度，Herdan 经过计算，确定为－2.4 级，然后再计算其理论分布，它和实际的调查几乎是一致的。

（b）除了分布模型外，对语料库的分析和统计通常还采取了一些统计量，如频数、散布系数、使用度等，我国的语料库统计通常采取了 Juilland 和 Chang-Rodisiguez 在计算西班牙文词汇频率时所提出的公式（孙建一 1989），Carroll 等在 AHI 语料库的统计和分析中，又提出标准词频指数（SFI，Standard Frequency Index），下面着重介绍他们所使用的公式：

频数（F）——单词在整个语料库中出现的数目，AHI 共收集了 5，088，721 个标形，在 70 年代初期，算是比较大的英语语料库。

散布系数（D）——这个系数在 0.000 和 1.000 之间，用作统计单词在不同题材的样本中的散布率。AHI 共有 17 个领域，如果一个单词只集中出现在一个领域，他的散布系数就是 0.000，如果在 17 个领域中都出现，系数就是 1.000。D 的计算公式是：

$$D=\left[\log P+\left(-\sum p_i \log p_i\right)/P\right]/\log n \qquad (7.4)$$

$-\sum p_i \log p_i$ 是最大熵，Carroll 使用了信息论的原理。

使用度（U）——在 100 万标形中的推算频数。U 是根据 D 来调整的频数，如果一个单词类型的 D 为 1.000，U 就是 F/5.088721，5.088721 是一个把实际频数调整到以 100 万为基础的频数，以便和 Brown 和其他语料库相比较。如果 D 小于 1.000，U 就会往下调整。如果 D = 0.000，U 就会取得一个反映该类型在 17 个领域的平均加权概率的最低值。U 除以 100 万就是相应的概率，例如一个单词的 U 为 1.2445，它的概率就是 0.0000012445。U 的计算公式是：

$$U=(1000000/N)\left[FD+(1-D)\right]f_{min} \qquad (7.5)$$

f_{min} 为 $\sum f_i t_i$，而 f_i 为单位 i 的频数，t_i 为每个领域的标形数目在总标形数中的比例。

标准频率指数（SFI）——它是 U 的对数转换，其公式是：

$$SFI=10\left(\log_{10}U+4\right) \qquad (7.6)$$

如一个单词的 SFI 为 90，这就意味着它在 10 个标形中会出现一次，80 为 10^2（100），70 为 10^3（1000），60 为 10^4（10000），依此类推。下面举几个例子，以见一斑：

表 7.12　AHI 的统计指标举隅

单词类型	F	D	U	SFI
the	373123	0.9969	73122.8	88.6
representative	53	0.8228	8.7265	49.4
instrumental	53	0.2012	2.4799	43.9
thankfully	1	0.0000	0.0022	13.5

the 是频数最高的，在全部标形中的出现率为 373123/5088721=0.073，representative 和 instrumental 的频数是一样的，但 instrumental 的 D 低于 reprentative 的，因为前者出现在 13 个领域，而后者只出现在 4 个领域，而且光在音乐书中就出现 43 次，故 U 也随着降低。thankfully 只出现一次，故 U 甚低。SFI 是反映这些单词的综合指数，实际上是反映以 100 万为基础的理论频数。the 接近 90，即 10 个表形中差不多有 1 次，而 representative 接近 50，即 10000 个表形中差不多有一次。其他以此类推。

(c) 逐词索引（concordancer）。作为一种研究的手段，我们必须建立一种快速而又方便的语料检索程序，这虽然不是语料库本身固有的，但却是设计语料库时所必须考虑的问题。目前所编制的计算机程序，除了生成各种词频表和给出各种统计量外，一般都有逐词索引的程序，有的程序甚至可以规定一定环境来逐词索引，例如把有搭配关系的词组（如 look up，look ... up）进行检索。下面是使用 Longman 的逐词索引程序检索 in 的部分结果：

表 7.13 **Concordance for "in" (96 lines)**

Text：NEWS. TXT	
15 e of American hostages from Iran	in 1979—80, stepped in at the personal
15 ges from Iran in 1979—80, stepped	in at the personal request of Mr Bush a
20 now believe that a breakthrough	in the hostage crisis is possible, becau
21 Hashemi Rafsanjani has been sworn	in as president of Iran. \ \ Al—Khale
33 ni is a rational man who believed	in dialogue. Under his rule Iran will c
34 s rule Iran will change its style	in dealing with the West. "\ * At lea
37 people were killed and 82 wounded	in vicious, all—night shelling between
39 ho targeted residential districts	in and around Beirut, police reported t
49 MAURICE Colbourne—Tom Howard	in the top TV soap series "Howards' Wa"
56 Colbourne 49, collapsed and died	in his wife's arms at their holiday hi
58 arms at their holiday hideaway	in France, just hours after a sailing t
65 s Jan Harvey, who plays his wife	in the series, said on behalf of the ca
91 psed with chest pains. She tried	in vain to save him, said his agent, S
100 newest series due to be screened	in the autumn. \Shocked \ "Howa"

上下文的长度是可以定义的，而且用户可以随时把那一行的全部上下文都检索出来。

(d) 语法标记 (grammatical tagging)。语料库汇集了大量语料，但是这些语料都是未经加工的原始资料，要利用这些资料作进一步研究，如识别语篇、制作语法分析器、进行机器翻译，等等，就必须对语料作语法标记，即把每个词所属的词类标出来。进行语法标记有两种方法：一种是以知识为基础的方法，即人工智能的方法，这种方法假定计算机要处理自然语言，就必须具有广泛的知识和逻辑推理的能力，即在知识的基础上进行推理。这是因为人类处理语言时的心理过程就是这样的，计算机就必须具有和人类知识相同的知识才能处理语言。另一种方法是以语料库为基础的方法，即概率的方法。这种方法并不认为计算机能够像人那样去解释语言，但是却认为人类处理语言时所采取的方法更像这种方法，而不像正统的人工智能的方法。它的假定是，如果我们对数量很大的语言数据作定量分析，我们就能进行概率性的预测，这就可以补偿计算机缺乏知识和推理能力的缺点。概率的方法也有其自身的缺点，它不能做到 100%的正确；而且它和一些流行的语言学观点相违背，例如 Chomsky (1957) 就指出，"一个人产生和辨认语法话语的能力并非建筑在统计逼近等概念的基础上的。"。但是英国语言学家 Leech，Sampson，Garside 等使用概率的方法作语法标记取得很大的成功，使世人瞩目。

他们编制了一个称为 CLAWS (Constituent-Likelihood Automatic Word-tagging System)，对 LOB 语料库的 100 万词进行标记，准确率达 96—97%。现在正在编制第二代的系统 CLAWS2，以进一步提高准确率。这个系统有 5 个步骤：

(a) 预编辑阶段。用半自动、半人工的方法为标记系统准备语篇。

(b) 标记赋值阶段。对输入语篇的每一个词都赋予一个或多个标记。这个阶段的赋值并不考虑语境，因此所赋予的是一套在各种语境中可能出现的标记。

(c) 熟语赋值阶段。程序需要寻找一些特定的单词或标记型式；在这些型式里，有限的语

境可以缩小可能赋予一个单词的标记的范围。

(d) 确立标记阶段。程序考察一个单词被赋予不同标记的各种情况，通过语境来确定一个标记，并且估计特定标记序列的概率。

(e) 后编辑阶段。对计算机的错误的标记作出人工处理，并重新格式化，以减少标记系统所产生的一些无关重要的信息。

在整个过程中，最关键的步骤是使用概率的方法来选择标记。Marshall (1987) 举了一个例子，假定我们要对 Henry likes stews 这句话进行标记。Henry 好办，它是一个名词短语 (NP)，我们可以提出两条语境框架规则来处理这个句子：

? VBZ——not NNS

NNS ? ——not VBZ

?是要进行标记的词。第一条规则是在动词第三人称单数现在时前不可能是动词复数。第二条规则是在名词复数后的不可能是动词第三人称单数现在时。但是 likes 和 stews 两个词都可以是动词第三人称单数现在时 (VBZ) 和名词复数 (NNS)，两条规则就解决不了问题。

CLAWS 使用的是另一种方法，它根据已经作了标记的 Brown 语料库，对一起出现的标记作统计分析，整理出一个 133×133 的过渡矩阵 (transition matrix)[①]，标出 133 种标记中的每一种标记在后一种标记出现的概率。再以此为基础找出概率最大的搭配，例如，

Henry	NP
likes	NNS VBZ
stews	NNS VBZ

就有

1 val (NP－NNS－NNZ－.) ＝ 17 × 5 ×135＝11475

2 val (NP－NNS－VBZ－.) ＝17 ×1×37 ＝629

3 val (NP－VBZ－NNS－.) ＝7×28×135 ＝16460

4 val (N－VBZ－VBZ－.) ＝7×0× 37 ＝0

决定其概率的公式为：

$$prob(x_i) = \frac{val(X_i)}{\sum_{i=1}^{n} val(X_i)} \tag{7.7}$$

概率最高的是第三种搭配 26460/11475＋629＋26460＋0 ＝ 69%，故程序确定这三个词的标记为名词＋动词第三人称单数现在时＋名词复数。

应该指出的是，提出概率的方法的人并不认为这个方法可以取代人工智能的方法，而是认为这两种方法都应得到支持。这是因为人们在使用语言的时候，既依赖规则，又依赖策略 (桂诗春 1995)，人工智能的方法基本上依赖规则，而概率的方法基本上依赖策略，把这两种方法结合起来，才能解决自然语言处理的问题。就目前的情况而言，人工智能的方法适合于一些要求较精确的，而范围较小的语言处理，例如某些类型的机器翻译，自然语言的前期处

① 表示过渡概率的矩阵，过渡概率是在前一个单位的条件下后一个单位出现的概率，例如在英语里，在 q 后面出现 u 的过渡概率是非常高的。

理（如数据库中的问答系统）；而概率的方法适合于其目的在于使计算机系统接受某一自然语言的任何输入的人机接口，如言语识别（对口语进行解码）、语篇转换为言语的系统（将书面文字转换为口语）、光学字母阅读（将印刷和手写的语篇读入计算机）、语篇批评（从正确率和语体接受性方面评估机读语篇）等等。

4. 语料库的使用。语料库的建立为各种语言研究和语言工程提供了很大的方便。

(a) 辞典编纂学进入了一个新阶段，语料库在收集例句方面取代了人工操作，使一些辞典出现了新貌。在英国，第二版牛津英语大辞典不但利用语料库增加了大量例句，而且使用了语法标记来帮助整理词头、发音、变体、引文、引文时间和交叉检索。它从 80 年代开始就通过数据库来收集引文，每年增加 12 万条。最后还出版了电子版牛津英语大辞典。Collins 的 Cobuild 英语辞典是在 2000 万词条的现代英语语料库的基础上编辑的，从选词、释义、举例等方面都利用了语料库所提供的资料。Longman 的科学用语辞典根据自己建立的数据库把基本的科学用语按概念组织编排成自成一格的同义辞典。作为辞典编纂未来方向的在线辞典编纂（on-line dictionary editing）也吸引了人们的注意，所谓在线编辑，实际上就是通过直接提取计算机文件资料，在计算机屏幕前编辑辞典。加拿大 Toronto 大学所编辑的《古英语辞典》(Dictionary of Old English) 就是用这个办法编出来的。

(b) 基本词汇表的编制，例如我国的《现代汉语常用字表》经过大量资料的统计，最后得出 3500 个汉字，其中最常用的 2500 字，次常用的 1000 字。为了计算机之间进行信息交换的汉字交换码也是在参考了静态和动态的资料基础上定出，《信息交换用汉字编码字符集・基本集》收汉字 6763 个，第一级汉字有 3755 个，第二级有 3008 个。我国大学英语教学大纲规定基础阶段结束后学生应掌握词汇量为 3800—4000 词，也是根据上海交通大学所作的 100 万字科技英语语料库，参考学生结束学习的抽样调查数据和对历届毕业生的社会需要凋查，而最后定出的。其他的一些外语词汇表虽然不是直接根据某一个语料库的资料来制定，但也参考了语料库的常用词词频排列而制定的。

(c) 大大促进了机器翻译和自然语言处理的发展，特别是人工智能方法和概率的方法相结合，展示了广阔的前景。Sampson 等人 (1987) 进一步提出用概率的方法来建立语法分析器。人工智能型的计算机语言 Prolog 和专门处理字符串的计算机语言 snobol，spitbol 等也走向成熟，处理语言的功能越来越强。

(d) 为计算机在人文科学的应用开拓了道路，使语篇和语体的定量研究更为精确可靠。以北美为基地的计算机与人文科学学会（ACH，Association for Computers and Humanities）已有 17 年历史，出版有双月刊《计算机与人文科学》，它和欧洲的文学和语言学计算机学会（ALLC，Association for Literary and Linguistic Computing）合作每年轮流在北美和欧洲召开世界性大会交流信息。研究的范围从各种语体和作家风格延伸到"造的不好的"（ill-formed）语篇和有错误的语篇（如学生英语）。由于采取了计算机进行统计，对语篇的观察更为精细，包括了词长、句长、作家爱用的词项和语法结构、语法类型（如名词/动词比例、动词/形容词比例），等等。Farringdon 等 (1978) 使用语料统计方法证明《查尔斯十二世战争史》的译者为英国 18 世纪文豪 Fielding 的研究，就是一个很好的说明。长期以来，人们一直怀疑 Fielding 是这本书的匿名译者，也发现一些根据，例如 Fielding 在他的作品中爱用的 hath、doth，durst 等词也见于这本译著。但是这种观察并非定量的。Farringdon 等用抽样的

方法建立一个包括了 Fielding 两本小说和 7 篇论著的语料库，另外又输入 6 种他的同代人的著作和《查尔斯十二世战争史》，然后对 Fielding 常用的短语、连接性短语、句子开始的短语、单词、句型进行比较，结果如下：

表 7.14　　Fielding 使用词语的比较（绝对数）

	Joseph Andrews	Fielding 其他短篇作品	非 Fielding 的作品	Smollet 模拟 Fielding 的讽刺作	查尔斯第十二世
A. 短语					
I beliveve /believe me	33	6	4	0	1
on/of a sudden	6	3	0	0	1
(all of/on a sudden)	0	0	1	2	3
in reality	11	4	0	2	9
by no means	10	2	3	0	6
in the world	25	3	1	0	9
叙述性、连接性短语					
and now	15	3	1	1	1
at length	28	2	9	0	27
upon which	12	8	5	0	23
C. 句子开头					
As soon as	16	2	0	0	38
As to the	9	14	0	0	61
At last	13	0	1	0	15
However	18	1	2	0	23
Now/And now	43	10	7	1	2
That	22	4	2	1	103
Upon this	6	1	0	0	13
D. 单词					
likewise	68	2	13	0	235
pray	27	1	1	1	0
E. 动词词组					
snapt/snapped his/her/my finger	9	0	0	1	0
readily/argee /accept	10	2	0	0	0
think proper to	17	3	2	0	43
do/did—verb	143	46	45	3	162
verb+not (inversion)	39	15	6	0	46
100 词中的总词数	124.6	35.0	65.8	4.5	244.8

根据上述资料的比较，Farringdon 等认为，可以断定《查尔斯十二世战争史》确实是 Fielding

所译的。Hope（1994）还进一步把语料库方法和社会语言学的研究方法结合起来研究莎士比亚一些有争议的剧作的作者归属问题，也取得有意义的成果。例如在早期现代英语（1500—1700）里，语言急遽变化，有些语言形式同时共存，如第二人称代词 thou 和 you，第三人称单数现在时 hath 和 has，助动词 do，等等。但是它们随着时间的推移而采取 S 弧形的变化，如（图 7.5）。

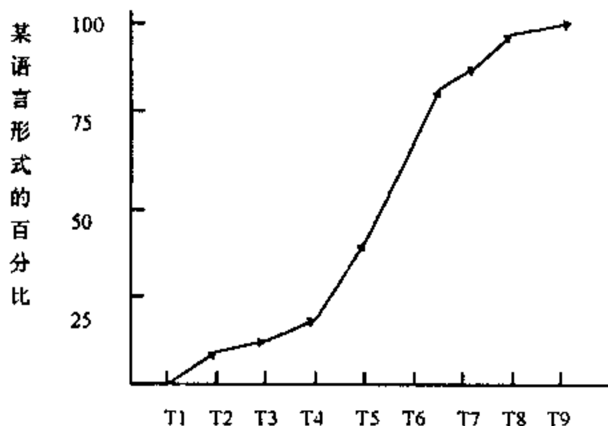


图 7.5 某语言形式的百分比

例如在 13 世纪，you 的用法甚少，到了 16 和 17 世纪有了迅速的发展，最后在 18 世纪才稳定下来，而 thou 的用法只保留在诗歌和宗教仪式里。更为重要的问题的，在两种语言形式共存的时候，早期现代英语的说话人使用 thou 的频率决定于几个因素，更为年青的、受教育程度更高的、更靠近城市的人更多地使用像 you 那样的新兴的、富有威望的变体。这些语言用法在人生中较早地定型，所以出生年月不同的两个作家在同一年代写作时使采用 thou 的频率也就不同。这也可以说，社会语言因素在形成语言变异型式中的起了作用，这些因素影响了早期现代英语作家对某些早期现代英语变量的使用。例如 John Fletcher 于 1579 出生在英国的东南部的上层阶级家庭，父于 1594 起出任伦敦主教，于 1589 后一直居住在首都。叔父为伊利莎白王朝的著名外交家。Fletcher 本人很可能是在剑桥大学受过教育。而莎士比亚和他的情况很不同，莎比他早生 15 年，生活在西南部中原地带的农村，而且出生微寒，没有受过高等教育。这些因素使 Fletcher 会比莎士比亚使用更多的新兴的、有威望的语言变体。因此我们可以通过这些变体（语言标记）来研究像《亨利第八》和《两个高贵的亲人》那样的交叉参与写作的作品里① 哪些部分是属于 Fletcher 或莎士比亚的手笔。

Hope 分析了助动词 do 在早期现代英语的用法，如（表 7.15）。在早期现代英语里助动词 do 是有选择性的，规则化的用法是 1500—1600 间兴起的一种变体。在 1400 年代，没有什么人使用；在 1500—1600 间，见于各种句子类型中；到了 1700 则已稳定下来成为规范。Fletcher 的年纪较轻，出生于较高阶层，受过高等教育，而且生活在城市，因此使用规则化的句子的场合应该比莎士比亚要多。Hope 统计了属于 6 个剧作家的 43 个作品的使用规则化句子的比例：按起出生年份，除莎士比亚（1564）外，还有 Marlowe（1564），Dekker（1572），Fletcher（1579），Middleton（1580）和 Massinger（1583）。（图 7.6）表示的每一本作品的使用规则化句子的比例，如图中的第一个条形表示在莎士比亚作品中有三部作品的使用百分比为 79%。值得注意的是他的用法没有和别的作家重叠，别的作家都没有低于 85%的，而莎士比亚的却没有超出 84%的。而且别的作者都有和另外的一些作者在使用上重叠的地方。这说明

① 一般认为 Fletcher 曾参与莎士比亚的《亨利第八》的写作，而莎士比亚也曾参与 Fletcher 的《两个高贵的亲人》的写作。

起码就莎士比亚而言，助动词可以作为一个强有力的语言标记。而 Fletcher 的用法则刚刚相反，没有低于 90%的。Hope 还用同样的方法来观察作家的出生年份和使用规则化句子的关系，使用的百分比是从莎士比亚（83%）→Marlowe（88%）→Dekker（89%）→Middleton（91%）→Massinger（91%）→Fletcher（93%），和出生年份的先后基本上是一致的。只有 Fletcher 的出生年份比 Middleton 和 Massinger 晚，但用得比他们多。Hope 用这个语言标记来观察《亨利第八》，说明它确实是莎士比亚与 Fletcher 合作的成果。我们这里着重的是介绍他所使用的研究方法，具体的结果就不再罗列了。

表 7.15

1. 规则化句子（现成为规范用法）	
(a) 肯定陈述句：	I went home
(b) 否定陈述句：	I did not go home
(c) 肯定疑问句：	Did you go home?
(d) 否定疑问句：	Did you not go home? / Didn't you go home?
2. 未规则化句子（现为不规范用法）	
(a) 肯定陈述句：	I did go home
(b) 否定陈述句：	I went not home
(c) 肯定疑问句：	Went you home?
(d) 否定疑问句：	Went you not home?

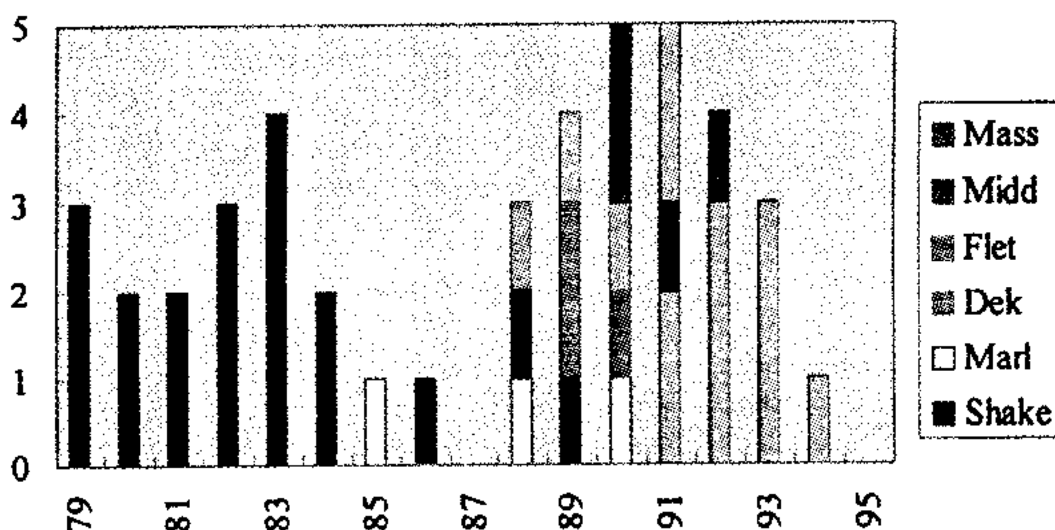


图 7.6　6 个剧作家的 43 个作品的使用规则化句子的比例

以研究语料库的开发和应用为目的的语料库语言学（corpus linguistics）已经成为一门羽毛丰满的学科。

8. 社会语言学研究方法

8.1. 结构语言学和社会语言学

结构语言学本来脱胎自人类语言学，而人类语言学所研究的一个主题是语言和社会的关系。但是结构主义强调的是语言形式，而且认为只要坚持其分析方法，语言学家就能取得某一语言及其社区的结论性的看法。结构语言学关心的是语言的普遍规则，对语言差异不感兴趣。但是语言差异和语言共同性是一个问题的两个方面，忽略了差异也会影响普遍规律的形成。Bloomfield（1927）深有体会地举了一个例子，他谈到一个作为资料提供人的 Menomini 族人，年约四十，英语不懂得几句，但他的 Menomini 语也糟透了，词汇量很小，曲折变化凌乱，只会用几个简单句型。他可以说是讲不出几句像样的话。但是这种情况不是个别的，可能代表了一代人。假定我们归纳他的情况，就会达到这样的结论：Menomini 语是一种没有人能讲出几句过得去的话的语言，因为无从比较，就索性说 Menomini 语词汇量很小，只有几个简单句型。这个例子说明，忽略了语言差异，也难归纳语言的共同性。Bloomfield 的看法是这种情况可能是英语战胜 Menomini 语所造成的影响。但是 Hymes（1974）举这个例子来说明，"言语社区在它们的正常的历史过程中，完全有可能发展它们自己在程度和方向上有所差异的语言手段，并在它们的交际生活中对这些手段赋予一定的地位。""总之，语言使用者的语言能力，乃至其语言本身，会发生变化，不过这些差异不会反映在通常描写范围内的语言结构里。通常描写的形式化的语言系统可能仅是不同的社会语言系统的一部分，这些系统的性质不能是假设的，必须加以观察。"所以在上述例子里，我们不能只研究 Menomini 语，而必须研究那个包括 Menomini 语、英语和其他语言在内的言语社区。

Hymes 用（表 8.1）来归纳从 30 年代到第二次世界大战后的结构主义所强调的方面：

表 8.1　　结构主义对语言结构和语言功能的看法

	所描写的	所比较的
语言结构	不变式	差异
语言使用（功能）	差异	不变式

这个表的含义是：就一种语言的描写而言，结构主义感兴趣的是找出其不变式，即同质性结构。就语言的比较而言，结构主义感兴趣的是找出语言结构的差异，有的人还认为差异越大越好。至于语言使用（言语）则是不大受到注意的，因为差异性甚大，仅能作为不变式的背景。如果我们要比较，不同语言在功能上是等值的，语言功能都表示为不变式。

Hymes 认为重点要有所转移，在（表 8.2）里，我们在一种语言的描写时，虽然也要考虑到不变式，但更应注意言讲社区的复杂性所产生的各种语言变体以及其相互关系。至于不同的语言的比较，由于语言类型学和语言共同性的开展，我们更应强调找出其不变式。至于语言的使用和功能，就一种语言的描写而言，我们应该找寻不变式，即社会语言型式。就不同

语言的比较而言，则应找寻语言使用和功能的差异。Hymes 还专门列表来表示“结构”语言学和“功能”语言学的重心是有所不同的：

表 8.2　　社会语言学对语言结构和语言功能的看法

	所描写的	所比较的
语言结构	差异	不变式
语言使用（功能）	不变式	差异

表 8.3　　“结构”语言学与“功能”语言学的重心比较

“结构”	“功能”
1. 语言结构（语码）作为语法	言语（行为、事件）结构作为说话方式
2. 语言使用对语码分析起执行、限制、相关的作用；语码分析先于使用分析	使用分析先于语码分析；使用语言的组织显示更多的特征和关系；表明语码和使用语码的有机的（辨证的）关系
3. 指称功能把语言使用规范化，完全语义化	全部语体和社会功能
4.（从跨文化或历史的角度看）要素和结构的分析是任意性的，或（从理论的角度看）要素和结构是有共同性的	要素和结构是文化上适宜的（或用 Sapir 的话来说，“精神上”适宜的）
5. 所有语言在功能上（适应性上）等值；所有语言在主要方面（潜能上）平等	所有语言、变体、语体在功能上（适应性上）有差异；它们在生存上（实际上）不一定等值
6. 单一的、同质的语码和社区（“一致性的重复”）	言语社区作为语码库或言语语体的一个矩阵（“差异性的组织”）
7. 像言语社区、言语行为、流利的说话人、语言和言语功能这些基本概念是想当然的、任意假定的	这些基本概念都是有问题的，需要考察的

社会语言学研究的是语言和社会的关系，有广义和狭义之分。广义社会语言学从社会的角度看语言，研究的是言语社区、多语制、语言态度、语言选择、语言计划和标准化、语言和文化等等问题；狭义社会语言学从语言的角度看社会，研究的是语言事件、语言功能、语码变异、语用、语篇分析、语言和性别等等问题。Hymes 把狭义社会语言学叫做话语文化学（the ethnography of speaking）。社会语言学开拓了语言学的视野，把研究的对象延伸到句子以外的领域，有的社会语言学家（如 Fishman，Labov）认为社会语言学促使了面向语境和功能的语言学的产生，因此它才是“真正的语言学”（real linguistics）。但是一旦社会语言学取代了语言学，社会语言学也不再存在了，因此这可以说是“一个自我消失的预言”（a self-liquiding prophecy）（Fishman（1971）语）。有的人则认为，虽然社会语言学强调语言的社会性，有其充分的理由，但它不一定要代替语言学，因为它还没有一整套成熟的语言理论。

不管怎样，社会语言学的出现孕育着更多的更新的研究方法，所以社会语言学对语言学的发展在某种程度上是方法论的发展。这是我们本章要考察的，总的说来，社会语言学的研究方法继承了人类语言学和结构语言学的传统，但又发展起它自身的一套方法。

社会语言学和人类语言学有着深厚的血缘关系，它所采用的方法很多都来自人类语言学，如自然观察、抽样调查、访问资料提供人、收集和记录资料、分类整理、归纳型式等等。但

是社会语言学研究的重心是在社会中不同的人群使用的语言的差异，它认为语言范畴和社会范畴是两个独立的，但有相互联系的系统，必须寻找它们两者之间的相互关系。它的假设是：如果我们分别测量这两套变量，我们就能了解社会结构和语言结构的系统变化。在计算它们的相互关系时，社会范畴通常是自变量，而语言变量则是依变量。这种方法可以称为相关分析法（correlative approach）[①]。但是有些社会学家（如 Garfinkel）特别是社会语言学家 Gumperz 则反对使用这种方法，Gumperz（1967）特别指出，相关分析对言语行为差异的解释很不充分，它首先并没有解释言语行为在不同社会里为什么会视社会范畴之不同而有所差异，其次是相关并不能使我们了解支配被调查人的实际交际行为的社会规范和规则，也不能了解他们对社会关系的感知差异。他提出一种交互作用分析法（interactionist approach）[②] 以取而代之。这种方法不主张平行地观察两个变量之间的关系，而是认为只有通过语言数据才能获得关于社会范畴的信息，因为社会测量总是牵涉到资料提供人和调查者本人对社会范畴的认识。"正如词义总是受到上下文的影响一样，社会范畴也必须从环境制约的角度去解释，像状况和地位那样的概念并非说话人的永久性特征：它们像音素和词素一样，已经成为抽象的交际符号。它们在分析家的抽象模型里也可以被分离出来，但是人们总是在具体的上下文中认识它们的，因此语言范畴和社会范畴的界限就模糊不清了。"

交际过程是一个统一的过程，说话人在整个过程中根据他们自己的文化背景对外部环境的刺激进行修正，并从刺激中产生有关环境的交际规范。（图 8.1）表示了规范怎样影响言语信号的选择。

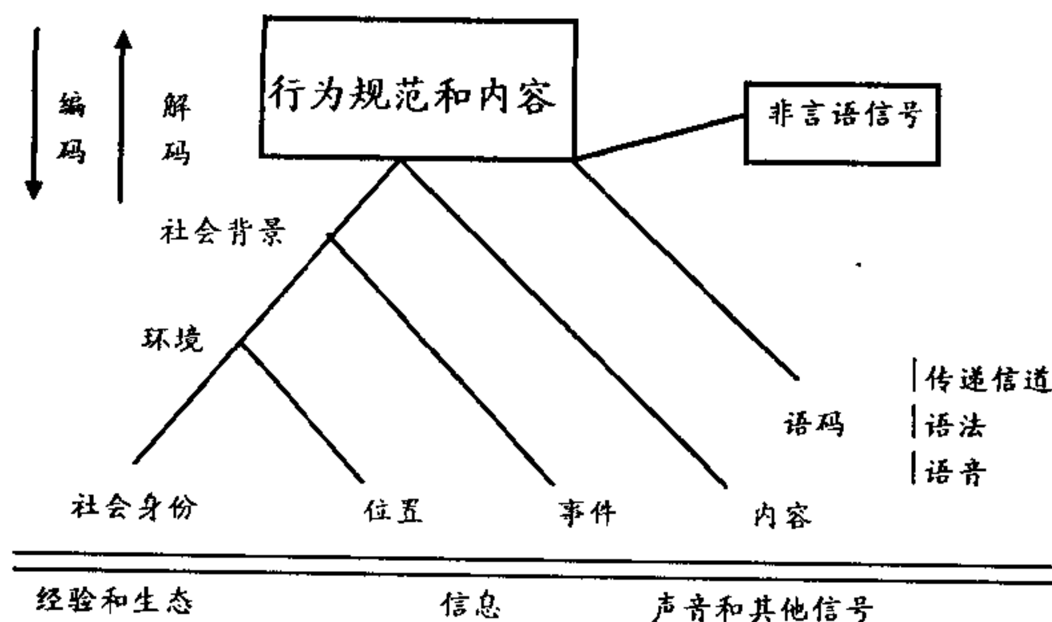


图 8.1 行为规范对语言选择的影响

首先，底部的双线把实际的交际过程和外部刺激区分开来，这些刺激被认知范畴转变为

① 相关研究的基础是统计学中的相关分析，请参看 11.5

② Gumperz 所主张的实际上是和定量研究相对立的定性研究方法。一般看法是定性研究应该和定量研究相结合。

交际符号。其次，经过范畴化后，说话人所要传递的信息都要被社会和环境的制约重新改造（树结构图），这些制约决定了说话人的社会身份，他在某些环境的某些社会事件中的权力和义务。第三，这些过程导致行为规范，并从中产生出合适的语码（言语变体），语码的选择进一步决定了传递信道（口头的或书面的）以及产生实际话语的语音和句法规则。

从上述讨论可见，广义社会语言学倾向于使用相关分析法，而狭义社会语言学则倾向于使用交互作用分析法。Gumperz 和 Hymes（1986）又把狭义社会语言学称为交际文化学（ethnography of communication），并且编了一本文集《社会语言学方向》，收集了使用这种分析法所作各种调查和研究。

社会语言学是在后结构主义时代出现的。在语言描写方面，它既注意定性分析，又力图作形式化处理，例如 Labov（1970）在描写黑人英语的否定式时，就使用了转换语法的描写方法。他认为，“就语言结构的共时研究而言，过于强调社会因素的重要性是错误的，转换语法虽然完全否认语言的社会环境，但在找出语言结构的不变关系时作出了重大的进步。现在已经很清楚，如果我们不认真研究促使语言进化的社会因素，我们就不可能在了解语言变化的机制方面迈出任何重要步伐。”可见 Labov 一方面维护其社会语言学家的立场，另一方面又不放弃使用语言学的一些新描写方法。

社会语言学不但强调收集数据，而且更强调数据的精确和可信，所以越来越多地采用定量的方法和统计的方法。Fasold（1984）指出，语言学研究只注意语言结构，而语言结构又比较有规律性，因此使用定性方法就足够了，无需用到统计学。可是当研究的对象是语言在语境中的使用时，我们就需要一种把真实和谬误区分开来的方法，而统计学为我们提供了这样一种方法。关于统计学的基本的原理和应用，我们将在下篇解释。但有些概念（如抽样）如果不讲清楚，也会影响我们对社会语言学的方法的了解，则在这一部分加以说明。

8.2. 社会语言学的抽样方法

抽样在社会调查中是不可缺少的，因为要调查的总体太大，不可能也没有必要对总体的每个成员都作调查。社会语言学[①] 抽样首先碰到的一个问题是资料提供人有很大的差异：他们不但有不同的方言特征，而且在不同的环境里有不同的说话方式和内容。Heath（1980）在调查英国 Staffordshire 的 Cannock 语时发现当地人说话时会不规则地拉下 h 音。如果我们只是在“理想的”操方言的说话人的个人语言（idiolet）中去调查语言变量，就会发现它捉摸不定的，只有使用周密的方法才能发现这些人际之间和个人语言之间的语言变量的一致性。

社会语言学既要研究社会的语言差异，就必须要有足够数量的、不同类型的语言数据，并且还要考虑收集语言数据的社会环境。“代表性”的概念应该扩充到不同说话人，甚至同一说话人所使用的不同类型的语言。Sankoff（1980）指出，要获取好数据意味着要在抽样程序中作出三种决策：

'1. 对样本的总体作出定义。

---

① 社会语言学所使用的抽样方法和定性研究方法中的抽样方法相似，基本上都属于有目的抽样，和定量研究方法的概率抽样不很一样。概率抽样的原理和方法详见 11.8。

2. 对社区内语言差异的各个有关方面进行估计，这牵涉到分层抽样。所以我们必须了解民族群体、性别、社会阶层对其所用的语言有些什么影响。多数研究表明它们和语言环境一样，在很大程度上都会对语言的使用发生影响。

3. 要决定样本数。

Sankoff 认为只要作出这三种决策，抽样就可以通过社会网络的方法或使用随机的方法而取得。

8.2.1. 对样本的总体作出定义

这就是对我们所感兴趣的群体或社区加以描绘，找出考察群体成员的样本框架。例如下面所介绍的 Labov 在纽约市的调查，排除了很多母语不是英语的人，这就牵涉到对母体的定义的问题。纽约市的移民很多，为什么要把他们排除在外？这是一个问题，另一个问题是怎样定义操本族语者？在社会语言学分析中，这往往也是一个引起争议的问题。凭他们说话的口音来断定其是否操本族语者，也很不容易。有的小孩的父母亲为移民，而他自己还是永远掌握不到他从小在其中生长的社区所使用的某些语言结构。Trudgill (1983) 指出，有些人在 Norwich 生活了一辈子，还是没有 Norwich 口音。

Horvath (1985) 在研究澳洲悉尼英语口语时，并没有排除把英语作为第二语言的说话人。这种做法有理论上的意义，因为少数民族的发音往往在语言变化中起到足以影响整个悉尼言语社区的主导作用。这种决策对研究方法也提出不同的要求，需要不同于 Labov 所采取的分析方法。至于英国所做的少数民族语言调查（The Linguistic Minorities Project，见 Milroy (1987)）更是直接考察在英国和威尔斯的非英语说话人所使用的语言的各个方面。这项研究的一部分是成人语言使用调查，它企图描写英国 3 个城市的 11 个少数民族语言群体，其目标和一般的城市方言调查不同：它旨在了解有哪些语言在英国使用，是怎样使用的。这项调查也面临着抽样的问题，调查者认为在 3 种情况下都会产生抽样的偏颇：(a) 采取了非随机的办法，它意味着有意识或无意识地受到人为的干扰；(b) 如果抽样框架（清单、索引或人口记录）并没有全面而准确地覆盖了总体；(c) 如果总体的一部分找不到或拒绝合作。实际上英国的移民群体是随机地分散在各地的，如果根据选民登记表之类的抽样框架 (sample frame) 来进行随机抽样，也不一定合适。调查者于是采用了两种方法：一种是少数民族姓名分析，从选民登记表上先找出少数民族姓名，然后再使用系统或随机的抽样办法；另一种是根据少数民族社区所提供的少数民族语言使用者的清单，例如在伦敦和考文垂便是这样找出意大利族说话人。之所以要依赖少数民族社区是因为他们不是英联邦公民，没有投票权，不会参加选民登记，而且意大利姓名又没有明确的区别特征。

这种建立抽样框架的方法也引起争议，但是这个调查的研究者认为，这种方法虽然难以作统计学评估，但在对总体的大小和地区均不知道的情况下，他们已经尽力保证抽样的代表性了。

8.2.2. 分层抽样

抽样的最基本的办法是随机抽样，即让总体的每个成员都有同样机会被选到。但是在实际的操

作中，往往会遇到很多困难，主要是抽样框架本身往往有偏颇，1936 美国的 Literary Digest 根据电话簿用户排列来作抽样调查，以预测当年总统后选人。它预测罗斯福会落选，但结果罗在 48 个州里的 46 个州里胜出。原因是当时使用电话的多为中上层阶级，而中上层阶级多为反对新政的共和党人。另外一个原因是社会中的变量很多，如果样本不够大，往往不能代表某些变量。为了保证样本能充分代表总体，一般采取分层抽样(stratified sampling)的办法。

Labov 在调查纽约市各个阶层的英语特征时，对说话人和他们的语言都使用了抽样的方法。它决定在纽约曼哈顿下东区展开调查，这个区人口不算多(1960 年的人口为 107,000)，但却是美国各民族各阶层的一个缩影。27%为犹太裔，11%为意大利裔，25%为其他的欧洲移民(包括乌克兰、俄罗斯、波兰、爱尔兰等)，26%为波多黎各裔，8%为美国黑人，3%为非白人(主要是华裔)。Labov 把他们分为四组：黑人、犹太人(正教徒和保守派)、天主教徒和基督教徒。在这个地区居住有不同的社会阶层，Labov 分为三个档次，下层(0－2)、中层(3－5)和上层(6－9)。在这个地区已经进行过一次广泛的调查，叫做“青年动员调查”(简称 MFY)，调查了 617 人。Labov 在这个基础上进行筛选，主要是去掉母语并非美国英语的人和已经搬走或死去的人。最后实际调查了 155 人，作为他的美国英语调查(简称 ALS)的样本。

表 8.4　　Labov 的 ALS 抽样方案

民族	阶层	MFY 总数	MFY 调查部分总数	母语为英语人数	搬走或死亡	ALS 目标样本	ALS 访问总数	ALS 语言访问数	ALS 电视访问数
(列)		1	2	3	4	5	6	7	8
黑人	0－2	32	32	29	6	23	16	14	2
	3－5	36	36	31	10	21	19	16	3
	6－9	17	17	16	11	5	5	5	—
小计		85	85	76	27	49	40	35	5
犹太	0－2	71	71	14	4	10	8	8	—
(正教	3－5	55	38	13	3	10	10	9	1
徒)	6－9	48	48	22	8	14	11	6	5
小计		174	157	49	15	34	29	23	6
犹太	0－2	25	25	9	3	6	4	3	1
(保	3－5	35	21	14	4	10	8	7	1
守)	6－9	40	40	35	13	22	18	12	4
小计		100	86	58	20	38	30	22	6
天主教	0－2	72	72	36	11	25	18	11	7
	3－5	102	69	41	17	24	19	15	4
	6－9	37	37	27	11	16	13	8	4
小计		211	178	104	39	65	50	34	15
基督教	0－2	7	7	4	0	4	4	3	1
	3－5	13	13	6	6	0	—	—	—
	6－9	27	27	15	10	5	5	5	—
小计		47	47	25	16	9	9	8	1
总数		617	533	312	117	195	158	122	33

Labov 的调查发现不同阶层的英语发音有差异，在后面（见 8. 4. 2）我们还要专门介绍的。这里只是讨论他抽样方法。总的来说，大家都承认他的程序在保证样本的代表性方面比过去的调查要科学和精密得多。他在两方面是有开创性的：一是他普遍地调查了各个阶层的说话人，而不限于某一阶层；二是他调查了他们的语体。但是从严格的统计学的角度来，很难说他的如此少的样本就能够反映总体的情况。而且他把母语非美国英语的人排除在外更引起争议。

8.2.3. 样本数目

从统计学的角度看，抽样数目多少决定于（a）总体的差异程度的大小；（b）容许误差的大小；（c）使用抽样方法。抽样数目太大会造成浪费，太少会使调查结果产生较大的误差，影响我们调查的结论。像 Labov 的调查不能说有统计的代表性，只能说是非专业的“代表性”。他所使用的实际上是 6.2.2 节里所谈到的定性方法中的有目的抽样，并非概率抽样。这在社会学调查里经常使用。Sankoff 指出，语言调查也许不像其他调查那样需要那么大的样本，因为语言行为比其他行为有较大的同质性。即使是调查较为复杂的社区，150 样本也足够了。但是这些样本必须选择得当，能代表我们所要概括的所有社会阶层。原因显而易见，在语言分析中对数据的处理有很多要求，因此样本数不能太大。Trudgill（1974）在英国诺威奇的语言调查只找了 60 人。Shuy 等人（1968）用严格的随机抽样的办法对美国底特律 254 个家庭作了 702 次访问，但是到了最后语言数据分析时，根据说话人各方面的适宜性，只选了 36 人。所以细心的抽样反而显得多余。

8.2.4. 判断抽样

判断抽样是有目的抽样的一个延伸，其基本原则是研究者事先决定他调查的说话人的类型，然后找出符合规定范畴的这类说话人的人数，再进行抽样。要使样本有效，判断抽样必须以某一个理论框架为基础，而且研究者能够表明他的判断是合理的、目的明确的。Romaine（1978）和 Reid（1978）对爱丁堡学生的语言研究就是根据 1971 年人口调查的资料，按照住房、教育、就业、健康等情况选择了处于贫穷地区和富庶地区的学校各一间，然后调查和比较不同社会阶层的孩子的语言情况。他们的目的明确，就是要观察规定得很清楚的社会群体的语言特征，因此没有必要再使用随机抽样程序。由此看来，要不要坚持随机抽样的原则应视情况而定，像少数民族语言调查，因为对总体的构成和特征所知甚微，只能用随机程序来保证样本的代表性。但如果是一个人口资料充分，而其特征又能明确规定的城市，判断抽样也是可取的。原因有二：一是语言调查的样本一般都不能作到概率抽样，勉强说它是概率抽样会招致很多学术上的批评；二是似乎相当小的样本（小到难以说有代表性）也能说明大城市语言差异。

判断抽样说明抽样方法往往和研究目的很有关系，视研究目的之不同，可以采取不同的抽样方法。Sankoff 建议采用社会网络的方法，这个方法认为，与其选取一些代表某一抽象的社会范畴的个人，还不如直接研究已经存在的社会群体。这样研究者就必须接近这个群体，和他们生活在一起，直接了解他们是怎样使用语言的。这个方法实际上继承了人类语言学的传

统，具有更强烈的定性方法的性质。

8.2.5. 语言的抽样

到目前为止，我们谈的都是怎样在社区中抽取说话人，但是社会语言学还有另一方面，就是在各种环境中抽取语言样本，即调查说话人在各种语言环境中所使用的不同语言类型。

众所周知，说话人在不同的场合会视乎其传递意义的目的和各种环境因素的不同而使用不同的语言类型。其中最重要的一个因素是说话人对他的对方的心理-社会态度，例如社会距离和亲密程度。有的语言有两种第二人称的用法（如汉语的“你”和“您”），使说话人可以对对方表示尊敬；在有的语言社区里，可以通过语言转换来表示同样的说话意义。Labov 调查单语制社区从正式的到随便的 5 种不同语体，就是一例。调查的目的是了解言语社区的规范以及说话人在不那么正式场合里转换语体的取向。

8.3. 数据的收集

样本选定后，接下来的问题是怎样从说话人身上取得数据，或是所取得的数据是否有效？这个问题之所以重要是因为语言使用型式对各种环境都很敏感，向说话人取得数据的方式会在很多方面影响我们所分析的数据。传统的方法是一对一的访问。Labov 建议访问者应该向每个说话人取得一到两小时的言语材料。但是很难作硬性的规定，如果是语音材料，二、三十分钟也能达到要求。但是如果我们想了解的是说话人使用某些语音变量的变化情况，就需要更长一点时间。如果想了解是句法使用，那就要更长的时间，因为有关的结构不一定像语音元素那么容易出现。由于数据只有经过分析后才能了解有没有取得所需的东西，所以我们还要和资料提供人保持联系，以便进一步获取资料。究竟需要多少数据决定于我们的研究目的：语音、词法、词汇和句法分析有不同的要求。

8.3.1. 数据的类型

Saville-Troike (1989) 指出，对每一项研究而言，并非所有的数据类型都是有用的，要进行社会语言学研究，应该注意取得下面的几种数据：

8.3.1.1. 背景信息

要了解一个社区的交际型式，就必须首先取得这个社区的历史背景资料，包括居地历史、人口来源、与别的群体接触的历史、影响语言和民族问题的重大事件。对调查地的一般的描述也是需要的，例如地形特征、重要标志的地点、人口移动的型式、就业情况、宗教信仰、学校的招生录取，等等。如果已经有公开发表的资料，也要收集。较新的资料则可从各级政府部门去收集。

8.3.1.2. 人工制品

社区的很多物体都有助于了解交际的型式，例如建筑物、符号以及像电话、无线电、书本、电视那样的通信工具。数据收集往往从观察这些事物开始，例如在访问中提出“这个东西用来干什么？”“你用什么来做……？”使用文化语义调查程序来对物体进行分类和称呼是了解言语社区怎样用语言来组织经验的第一步。

8.3.1.3. 社会组织

社区组织的各方面的情况也应注意收集，例如领导人和官员的办公室所在地、商业区和专业区的分布、权力和影响的来源、各种正式和非正式的机构、民族和阶级关系、社会阶层的组成和分布，等等。这些资料可从报纸、电台、政府公报中获得，或是通过对背景的样本的系统观察，对不同阶层的人民的访问来收集。也可以作社会网络分析，以了解什么人和什么人进行交往，以什么身份和为了什么目的而进行交往。也可以通过网络来了解一个异质性的社区中的各个子群的界线，并了解它们的长处在什么地方。

8.3.1.4. 法律信息

有关语言的法律和法庭裁决也是重要的数据，例如什么构成“诽谤”或“诲淫”，“言论自由”的界线是什么，例如西德法庭在1980年宣判两个纳粹党卫队队员无罪，“因为所有的证据都是口头的，并没有一项书面证据”。法律甚至规定政府文件使用什么语言，有的地方（如美国的选举法）还规定只要全民使用该语言超过5%，选举票上就应该印有该语言。在多语制社区，由于语言的沟通出问题而导致法律诉讼常有出现，这就要求法律家和语言学家共同协作，才能达到法律公正。于是法律语言学（forensic linguistics）就应运而生（见吴伟平(1994)）。与语言有关的法律信息一般都是正式语码化的，可以在法律书籍和法庭记录中找到，也可以通过访问各种法律事件的参与者或亲身到现场观察而获得。

8.3.1.5. 艺术数据

文学作品（口头和书面的）也很有价值，因为它们对社会现象给予艺术的表现，而且还反映了社会对语言的态度和价值观。在文学作品里的交际型式体现了某种社会规范，而且刻画了不同类型的人对语言的典型使用，例如戏剧作品提供了大量社会阶层使用语言的素材（如萧伯纳的《卖花女》）。艺术数据还包括歌曲、说书、相声，甚至书法。

8.3.1.6. 普通常识

语言使用和解释背后的一些假设很难看出来，因为它们往往表现为一些没有说出来的预设，但是它们以“众所周知……”和“正如大家所说的，……”的公式或以格言和警句的方式，使这些假设浮到表面。这些都是无须证明的“事实”，它们支配了各种交际行为。有些资料可以用这样的问题诱发出来：为什么在某一特定环境里你是这样说的，而不是那样说的？下面还要谈到的文化学（ethnoscience）和文化方法论（ethnomethodology）直接与发现这类数据有关。

8.3.1.7. 语言使用的信念

民俗学家长期以来都对这类数据，包括禁忌语和它的后果，感到兴趣。这些信念还包括谁或什么东西才能够说话，谁或什么东西可以与之交谈（上帝、动物、植物、死人，等等）。与此有关的还有关于语言的态度和价值观的资料，包括对沉默寡言与滔滔不绝的肯定或否定态度。

8.3.1.8. 关于语码的数据

在社会语言学里，把语言看成是一些静止的词汇、语音、语法单位是不适宜的。这些语言数据虽然也十分重要，但是应该把它们纳入更广阔的领域内来进行观察和研究。它们应该和副语言特征（paralinguistic）和非言语（non-verbal）特征一道，都是人们在语言事件中所使用的工具。如果我们要调查的语言并非自己的母语，我们还要阅读有关这些语言的词典和语法书籍，以便分析这些语言单位在社区中是如何使用的。

8.3.2. 数据收集程序

在一个言语社区内收集语言使用的资料，并没有单一的最佳方法。哪种方法较为合适要决定于调查者和社区的关系，所收集的数据的类型，调查现场所处的环境。现场调查程序的好坏主要是看我们能否避开记录者的偏颇性的感知，能否在自然环境下对交际行为进行观察。因此调查者应该各种调查方法均备于我，按照实际情况，决定使用那一种方法。一般来说，社会语言学调查继承了人类语言学传统，具有定性研究的性质，但是定量的方法在数据收集时也是十分有用的，特别是用它来测量行为一致性的程度和不同情况下语言差异的性质和数量。但是使用定量方法时，必须首先用定性方法的程序来支持其效度，而定量方法反过来又支持定性观察的信度。数据的效度是一个核心问题，根据无效的数据来形成的结论是站不住脚的。Labov（1972）根据过去现场调查的经验，总结了5条社会语言学方法论的公理：

1. 语体转换。没有人只用单一的语体，但有的人使用多一些，有的人少些。说话的环境和题目改变时，说话人的语体会随之转换。这些语体有特定的语言特征；有些转换可从说话人的自我改正中看出。

2. 注意。注意是定义不同语体的范畴，"不同语体可以沿着单一维度而排列，按其对言语的注意程度而测量。"（Labov）注意的作用最明显低表现在人们对自己的话语进行监察。例如我们观察可说话人在非正式场合和情绪激动时所使用同一语言变量的情况，这两种语体的相同特点是说话人对自己的言语给予最低限度的注意。

3. 本地话。语言学家对语体连续统上各种语体并非同样感兴趣，有的语体表现出不规则的语音和语法型式，带有很多"过度改正"。有的语体则比较系统，本地话是说话人最自然的交际手段，是他们最不注意自我监察的。对本地话的观察可以为分析自然言语行为提供最系统的数据。

4. 正式场合使用的语言。这是访问场合所使用的语言："对一个说话人的系统观察导致一个正式的环境，它对言语的注意远非最低限度的。"在了解和提供资料的访问主体里，不会使用本地话。虽然说话人可以尽量作到较随便和友好，我们还是可以假定他还有一种更为随便

的语体。

5. 好数据。要获得足够的好数据的唯一的方法是使用个别的录音访问，即通过最明显的系统观察。

这就产生所谓“观察者的悖论”(Observer's Paradox)：对社区语言考察的目的是了解人们在他们不受到系统观察的情况下如何说话的；但我们又只能通过系统观察来获得数据。(试和索胥尔悖论相比较：通过观察个人来研究语言的社会性，但研究个人的一面必须通过观察社会的环境里的语言)Labov 认为这些问题并非不可解决的，我们应该寻找一些补充正式场合访问的数据，或使用各种方法改变访问的结构。(见 8.3.2.3) 收集数据的方法很多，Saville-Troike 谈了 7 种，我们归纳为 5 种。

8.3.2.1. 内省法

内省法 (introspection) 不但是收集自己言语社区的数据的一种手段 (参阅 7.2)，而且也是为了收集数据而必须培养的一种技能。这种方法之所以重要是因为我们不是为收集数据而收集数据，而是因为每一个人都有自己的文化背景，要回答语言和文化各方面的问题必须从调查者和被调查者的角度来给出答案。如果调查者本人是双文化的，他就必须把两种文化的信念、价值、行为加以区别开来，这种比较对群体和个人都会提供有用的信息和启发。

在一个培训班里培养这种技能的最有效的手段是，让每一个人根据自己的经验回答关于交际的各种问题 (见 8.4)。然后把反映文化规范的答案和实际发生的真实情况加以比较，认识它们差异的意义。人类学家早就注意到“理想的”和“实际的”差别：这不是真和假的问题，更不能加以否定。实际上这是承不承认特定行为的问题。这可以比喻为开车人看见停车的红灯，“理想的”做法是把车停下来，而“实际的”特定行为是把车放慢 (而不是完全停顿)，有时还有人一点都不停下来。把“理想的”规范和“实际的”情况加以区别是客观地观察文化的重要阶段。对语言和文化问题的回应通常都是“理想的”答案，它们是群体成员的正规教育的一部分。“实际的”行为处于一个连续统，是从非正规模型中学回来的。它们通常都是在无意识中产生的，难以有意识地去认识。如果把一些“实际的”行为点出来，有人还会干脆否认它们曾经发生过，或认为这根本是无关大局的小事。

如果研究者对自己言语社区的语言型式有确切的了解，他就可以把自己的观察和别人的观察加以比较，把自己的观察和经过系统观察而获得的客观数据加以比较。

8.3.2.2. 观察法

观察法有多种 (见 6.2.1.2)，最常用的一种方法是参与性观察。研究者本人总是一个言语社区的成员，生来就具有这种身份；如果他能溶入另一个社区超过一年，他就有可能观察和了解其文化型式。关键在于他能否尽量摆脱他自己的文化经验的影响，这就需要观察者具有相对的文化观，了解不同文化的差异，对别的文化进行观察时能够具有敏感性和客观性。这就是从 Malinowski 开始的一次现场调查的革命。参与性观察的最大特点是调查者可以随时通过有意违反交际规则，通过观察和提取反应，来检验他所形成的交际规则的假设。参与群体活动必须有一段时间，重要的信息才能出现，和社区的成员才能建立起一种互相信任的关系。局外人的身份容易产生问题，如果一定要保持这样的身份，这个身份应该是主人承认和希望的，对群体的福利有贡献的，例如作为一个教师和建筑工人，换句话说，调查者必须使人感

到，他不是光在那里“提取”数据，而是作出回报，使社区得到好处。

采用参与性观察方法要求有高水平的语言和文化能力，如果观察的时间有限，这更是现场调查成功与否的必需的条件。作为一个参与活动的观察者，为了毫无阻碍地参加各种语言事件，并使别的参加者乐意与之交谈，调查者应和社区的其他成员一样，具有同样的语言背景和语言能力。在调查者亲自参与活动的情况下收集数据，还要包括调查者自己和别人的关系的数据，因为我们往往需要分析调查者在交往中所起的作用。

虽然参与性观察是主要的，但也可使用非参与性观察来收集数据。有些地方经专门设计，房间里配有单向镜子，可进行不惹人注意的观察；也有些地方可以容许调查者在场观察，而不致于影响进程。另外在观察一些像集会那样的动态过程，和社区成员关系不那么密切的观察者最好不要过于积极地参加活动。

使用录像办法来观察交际行为是参与性观察的一种很好辅助手段，因为录像可以重放，进行细致的分析。但是录像机的视野较小，必须把录制的材料放在一个整体的环境中去理解。不同社区对录像、录音、照相，甚至记笔记会有不同的接受程度，不能勉强。

8.3.2.3. 访问

访问在开放性社会里是司空见惯的，也是收集各种各样文化信息的常用手段，包括亲属称谓、宗教和社区重要事件、民间传说、历史故事、歌曲、专业知识说明、社交活动描写，等等。在访问中，我们使用提取法（elicitation method）有目的地收集语言数据。组织访问必须注意几个问题：

1. 选择可靠的资料提供人。那些在局外人看来最容易找到的人往往不是社区里有代表性的人物，因此可能提供不准确的、不完全的信息，和研究者从别的成员中所获得的信息冲突。

2. 提的问题必须是文化上合适的。调查者应该知道哪些提问是合适或不合适的，为什么？在哪些方面是合适的或不合适的？

3. 对交际中表示同意、不安、愤怒、讽刺等信号很敏感。这些敏感性有助于了解供资料提供人的可信程度和问题的适宜性，有助于决定什么时候应该结束访问。

4. 对数据要有标音、整理和分析处理，其结果和最初收集的数据会有些不一样，和研究者的理论方向也会有些不一样。如果访问者使用的不是他自己的本族语时，还需要精通另外一种语音字母，用它来进行标音。

如上所述，访问是一种适合于使用正式语体的语言事件，牵涉到陌生人双方的交往，但是说话人身份很明确。双方不是处于同等地位交谈，并没有同等的轮流讲话的权利。谈话的一方（访问者）控制着交谈的进程：他选定话题和提问的方式。被访问者由于同意访问，必须以合作态度回答问题。按照会话的合作原则，回答应是简短的、相关的。访问者得到回应后，应该提另一个问题。参与一个访问的语言事件的双方充分意识到他们的身份，所以轮流讲话的机会不是平等分配的。

Labov 指出，现场调查者应想办法克服访问的这种型式，以取得一种“有利的交谈地位”，因为访问中双方的不对称的地位往往使语篇的结构受到影响，例如访问者为获得信息而作的直接提问在日常生活里就不常见，而且不同社区对直接提问的含义还有不同的理解。Hymes (1972) 还注意到一些受文化所决定的提问行为：智利的土著 Araucanians 把重复提问看成是侮辱，巴西的土著 Cahinahua 认为对第一次提问的直接回答意味着没有时间交谈，一个模糊的

回答意味着在第二次提问时才会作出直接回答，意味着交谈能够继续下去。

要解决这个问题，可以采取几种办法：

1. 改变访问环境。采取各种方法使访问人从正式的语体转移到本地话，例如让他离开主题去谈论童年发生过的一些事情，Labov 认为“死亡危险”是一种成功的问题，“你有没有处于一种严重死亡危险的时候?”像这样的问题打断了正常的访问，使被访者转移到使用本地话。

2. 小组讨论。人们在和朋友自然交往中往往会使用“正常”言语行为。在小组讨论时，系统观察的影响可以降低到最低限度。

3. 快速和无名访问。在上两种访问中，被访者的身份是很明确的。我们也可以利用不能算是访问的会话来进行无名的系统观察，例如 Labov 在纽约第五大街找了 3 间代表了不同消费水平的百货商店，然后向售货员打听，“Excuse me, where are the women's shoes?”（对不起，卖妇女鞋的地方在哪里?）通常的回答是“Fourth floor”（在四楼）接着调查者装着听不清，又问道，“Excuse me?”得到更为小心的强调式：“*Fourth floor*”通过这样的办法，*Labov* 就能采集到在随便的和小心的两种语体中的/*r*/的发音资料。

4. 非系统观察。用来验证在访问、小组讨论中所获得的数据能否代表本地话。我们可以通过公共场合(如火车站、公共汽车站)的会话来检查一些语言变量，以纠正系统观察的结果。

5. 大众传播。从广播和电视中取得数据，说话人的地位和言语手段一般没有很大的变化，大多是使用正式语体。近年来有些在出事地点所进行现场采访，被访者受到刚刚发生的事件的影响，就不会注意监察自己的说话。

6. 语体变化。说话人视其对自己的言语的注意程度的不同而改变其语体。Labov 在纽约市的调查区分出 5 种语体：(a) 随便的语体（在家里和朋友间使用）；(b) 小心的语体（半正式的，通常使用于访问）；(c) 阅读体；(d) 词表；(e) 最小对立体。(c) 到 (e) 为有监察的阅读语体，其正式程度从 (c) 向 (e) 增加。

所以访问应该是开放性的，应该有很大的灵活性，尽量避免事先设定的框框，注意在访问过程中出现的新思想、新信息、新型式，注意“理想的”和“实际的”文化之间的差异。提问也应该是开放性的，避免提那些可以用事先设定的答案就能回答的问题。例如，关于方言和区别它们之间特征的问题：“某某地方的人的说话和您的说话有什么不同? 您能听得懂他们的说话吗? 您能否举些例子?”关于对语言变体的态度的问题：“谁讲得‘最好’?”“谁讲得‘可笑’?”“他们为什么那样说?”关于不同的语言事件的认同的问题：“他们在干什么?（指交往方式)”“这是什么人说的话?”关于言语中的社会标志的问题：“您怎样对年纪大的人问候? 对年轻人问候? 对您的老板问候?”

为了进行统计分析，也需要编制封闭性的问题。但是必须事先确定可能得到什么回答和可能作出的什么解释。但是封闭性的问题总是违反开放性原则。

应该注意的是，哪怕是最简单的问题，其答案往往也会有文化特征。例如对年龄和小孩数目的提问要很小心，不同社区对基本年龄有不同的看法，对“有多少个小孩?”也有不同的理解：有的社区只算活的，有的只算男的，有的只算答话人自己性别的小孩。有人在美国调查坦桑尼亚人，发现问“有多少个小孩?”是不合适的，因为被访人回答“我们不计算小孩数目。”所以他要问“你有多少个小孩生在坦桑尼亚? 有多少个生在美国?”才能得到他想要得到的信息。

在封闭性问题里，往往会使用一些分级量表（例如语义微分调查），在这些问题的前后最

好有一些开放性的问题，例如Saville-Troike先让来自不同国家的学生按照他们自己社区里的情况，排列一些代表男性或女性的典型特征，如有抱负、喜欢竞争、盛气凌人、有同情心、圆滑，等等。然后再利用这些答案来提问，以了解这些特征如何不同程度地反映在男性或女性的说话里。差不多所有的学生都把男性排在较盛气凌人那一个档次里，但有些人认为盛气凌人主要反映在男性的说话里，而另一些人则认为这个特征反映在男性的沉默寡言里。同样的，调查者让学生在"友好的"的量表上排列不同的社区成员以后，再让他们讨论什么是友好的言语行为，发现大家的理解很不相同，例如闲聊过多在西班牙学生看来是友好的表现，在日本学生看来则是不友好的，因为他们认为闲聊过多表示社会距离，而不是友善。一个日本学生甚至说，"如果你和一个人是很要好的朋友，你就没有必要和他谈那么多。"

8.3.2.4. 文化语义学

文化语义学（ethnosemantics）主要是从资料提供人的语言中提取抽象程度不同的各种词语，分析其语义结构，以了解经验是如何组成范畴的。一般是采用分类法或成分分析法。我们对一个言语社区的交际过程的考察往往离不开那些有文化特征的范畴和情景。Frake（1968）指出，调查者往往采用"找出代表事物的词语"的方法来了解词义，例如指着石头问，"这叫做什么？"如果回答是mbaba，那mbaba就是石头。这实际上是在两种语言之间寻找词语的一对一的对应关系，往往忽略了文化特征。他建议的方法是"找出词语所代表的事物，"我们应该按照所调查的社区成员的概念系统去定义物体，这就不是去把语言重新编码，而是去了解所调查的社区的"事物"。例如我们去了解的是食物，然后再根据该社区的认知系统去把食物进行分类：

表8.5　　根据一个社区的认知系统对事物分类

something to eat（食物）				
sandwich（夹心面包）		pie（馅饼）		pie-cream bar（雪条）
hamburger （汉堡包）	ham sandwich （夹火腿面包）	apple pie （苹果馅饼）	cherry pie （樱桃馅饼）	Eskimo pie （爱斯基摩雪条）
A	B	C	D	E

Eskimo pie和apple pie与cherry pie虽然都有一个pie，但它其实是指"雪条"而不是"馅饼"，故不能把它们归为一类。同样的例子在英语是不少的，例如说"Look at that oak，"oak指的是white oak（白栎），而不是poison oak（毒栎）；blackbird是一种鸟，而redcap则不是一种帽子，而是一种鸟。人类学家对亲属称谓所作的研究也表明各民族的亲属称谓有相同点，也有差异处，不能简单地寻找词语的对应。英语的brother，在汉语可能是"哥哥"，也可能是"弟弟"，因为汉语的亲属多了一条年龄的界限，以说话人自己的年龄为基准可区分为兄和弟，以说话人上一代的年龄为基准，父系也可区分为叔和伯，母系的可区分为大舅和小舅。这都说明比较要先从事物入手，再找出代表事物的词语。

可以有两种提问方法来收集文化语义的数据。第一种关于范畴的分层结构，如：首先问"有些什么侮辱行为？"回答为"有友善的侮辱和不友善的侮辱。"再问下去，就是"有些什么友善的侮辱？"和"有些什么非友善的侮辱？"可以逐步深入，以了解范畴的分类。第二种是

关于特征集合的，目的是了解说话人是怎样在一个维度上进行比较的，如问“这两个物体/行为/事件有些什么不同?”“它们有些什么相同处?”“在这三个物体中，哪两个更为相似？怎样相似法?”

这种描写方法的最终目的是使用对社区成员有意义的范畴来对数据作出 emic（内部的）的说明，用事先决定的范畴来对数据进行 etic（外部的）分类，有利于参考和比较，但不是描写的最终目的。

8.3.2.5. 文化方法论与交互作用分析

文化方法论（ethnomethodology）是 Garfinkel（1972）提出来的，它主要是研究说话人产生和解释交际经验的基本过程，包括那些没有说出来的种种关于共享知识的假设。Garfinkel 认为，社会知识是在交互作用中表现出来的，所以对交际过程的描写应该是动态的，而不是静态的。

文化方法论和文化语义学的共同点是两者都把阐明一种文化的成员的能力和知识放在首要的地位，但文化语义学主要是通过民间词语来了解知识，而文化方法论则采取更为广阔而有没有那么形式化的态度去考察交际行为，其兴趣在于那些作为交际行为（言语的和非言语的）基础的解释过程。对 Garfinkel 来说，意义是“定位的意义”（situated meaning），就是说，意义由言语活动参加者在特定的环境中建立并积极地作出解释的。因此方法论指的不仅是怎样研究，它本身就是研究内容的一个部分。

Garfinkel 认为，社会学家往往把“结果”和“过程”区分开来：“结果”指的是对一些实质性的事物取得共同的认识，而“过程”指的是人们为了了解一个人所说的和所做的是否符合规则而采取的各种方法。在会话分析中，要报告“结果”较容易，但要报告“解释过程”和定位的意义却不那么容易。Garfinkel 让他的学生把普通会话中实际说的话和它的解释过程写在一页纸的左边和右边，下面是一个学生报告他和自己妻子的对话：

表 8.6　　实际的对话和其解释过程

丈夫	丹纳今天不用抱起来，就能把一个硬币丢进停车计时表里。	今天下午我把我们4岁的儿子丹纳从幼儿园接回家，他居然不用我把他抱起来，就够得着把一个硬币丢进我们经常停车那个区的停车计时表里。
妻子	你把他带到唱片店里吗?	既然他把硬币丢进停车计时表，就说明他和你在一起的时候，你没有开车。我知道你不是去接他就是往回走时把车停在唱片店附近。你究竟是往回走时和他在一起，还是把车先停在那里，然后再去接他？还是把车停在什么别的地方?
丈夫	不，去了补鞋店。	没有，我在接他途中去了唱片店，回家路上和他一起去了补鞋店。
妻子	为什么?	我知道你去补鞋店的一个原因，但事实上你去干什么?
丈夫	我去给我的鞋买双鞋带。	你记得吗？我的一只棕色牛津牌的鞋子的鞋带那天断了，所以我去鞋店买鞋带。
妻子	你的平底便鞋也很需要换个新的鞋跟。	我想到的是你应该弄些别的什么。你应带上你的平底便鞋，它要换新鞋跟了。你应该快点把它弄好。

由此可见，说出来的和实际谈论到的是有差别的，说出来的是所谈论到的内容的一部分，往往是较为简单、省略、隐蔽，甚至是歧义的。所以文化的基础并非共享的知识，而是共享的解释规则。Garfinkel 把文化看成共享的解释规则，这一点和 Chomsky 的生成转换语法的深层结构相似。在 Chomsky 来说，每一个说话人都具有一种使他能够创造性使用语言的语言知识；在 Garfinkel 看来，每一个说话人都具有一种能够使他创造性使用语言的说话知识。

Gumperz（1977）指出，意义在会话交互作用中的传递是有共同规律的：

1. 意义和对说话方式的相互理解部分决定于环境和说话人以前的经验。

2. 意义是在交互作用过程中通过协商取得的，它决定于前面的话语的意图和对话语的解释。

3. 会话的参加者总是在进行某种解释。

4. 对目前所发生的事物的解释都可以从以后所发生的事物的角度加以倒述。

这里出现一个鲜明的概念：说话人必须和以后开展的任何深度的和任何时间的会话交谈共享经验。

Gumperz 根据这个思想建立一个关于社会知识如何在说话过程中被存储、提取以及如何和语法知识相结合的理论框架。会话推断是受环境制约的解释过程，会话的参加者靠它来估计别人的意图，并以此为基础作出自己的回应。不同的参加者因为不属于同一社区，有不一样的文化基础，对会话中意义会有不同的理解。跨文化的误解说明，交谈者对事物的含义和预设、非语言的环境、非语言的提示等等都是很不一样的。

8.4. 数据的描写和分析

8.4.1. 定性分析

定性方法的一个重要特点是寻找事物的型式，广义社会语言学中对多语制的研究就是采用了这样的方法。很多国家都有多语制，但有没有一些基本的形式？从 60 年代开始就有人致力去寻找能够反映世界各国的多语制的型式，使用了两种方法：类型学的方法和定性公式的方法。类型学的方法根据一些变量来建立足以区别不同多语制国家的范畴，强调的是国家和民族的历史发展、不同语言在其中的合法地位、占统治地位的民族在国家中的相对的作用、语言的发展和相互关系、说不同语言的人口比例，等等。定性分析则企图把不同的类型归结成为一些公式，以揭示一些社会语言学的事实。我们不妨看看 Ferguson（1966）所建立的公式，他首先建立了三个区分民族语言的范畴。一种是主要语言（Lmaj），满足下列三个条件之一，就是主要语言：

人口中有 25%或多于 100 万人把它作为民族语言而使用。
它是国家的官方语言。
它是 50%的中学所使用的教学语言。

另一种是小语言（Lmin），它不能满足主要语言的上述条件，但却具有下面的特征

之一：

人口中有5%或多于10万人把它作为民族语言而使用。
它在小学的头几年后被用作教学语言，并用来编写小学课本乃至其他教科书。

第三种范畴是不能满足上述条件，但仍有意义的语言，可称为特殊地位的语言（Lspec），例如宗教的语言，在中学里作为一门科目而教授的文学语言，或作为共同语（lingua franca）而使用的语言。

除了这三种范畴外，Ferguson 和 Stewart（1962）都提出一些类型（type）和功能（function）。Ferguson 提出5种语言类型：

本地语（Vernacular，V），一个语言社区的非标准化的民族语言。
标准语（Standard，S），标准化的民族语言。
古典语（Classical，C），曾经是作为民族语言的标准语。
混合语（pidgin，P），由一种语言的词汇和另一种或几种语言的语法结构混合组成的语言。
民族混合语（creole，K），成为一个语言社区的民族语言的混合语。

这些不同类型的语言可以有以下的一些功能（用小写体英文字母表示）：

群体功能（Group，g），主要用于某一个语言社区的交际，以区别某一个社会文化的群体。
官方用途（Official，o），法律上规定在全国范围内作为官方或政府使用的语言。
作为更广泛交际用途的语言（Language of wider communication，w），作为一个国家内民族之间交际用途的共同语。
教育用途（Educational，e），作为小学低年级以外使用的语言，并有用该语言编写的教科书。
宗教用途（Religious，r）
国际交际用途的语言（International，i），和各个民族进行交往所使用的语言。
学校科目功能（School-subject，s），作为学校一门科目、而并非作为授课语言而学习的语言。

Ferguson 认为，这3类信息可以结合起来形成一条表示乌拉圭的语言状况的公式：

$$3L = 2Lmaj\ (So,\ Vg) + 0Lmin + 1Lspec\ (C)$$

其文字的表达式为："乌拉圭有3种语言，两种为主要语言：一种是满足官方功能的标准语（西班牙语），一种是满足群体功能的本地语（Guarani），都不是小语言；但有一种特殊用途的语言：满足宗教功能的古典语言（拉丁语）。"

表 8.7　　按属性定义的语言类型

属	性				
标准化	独立性	历史性	生命力	语言类型	符号
+	+	+	+	标准语	S
+	+	+	−	古典语	C
+	+	−	−	人工语	A
−	+	+	+	本地语	V
−	−	+	+	方言	D
−	−	−	+	混合语	K
−	−	−	−	民族混合语	P

除了这个基本的标记系统外，Ferguson 还提出 3 种补充手段，例如用分号来表示语体的高低，所以阿拉伯语在摩洛哥的地位是 C；Vorw，意味着“在双语制中，阿拉伯古典语是高变体，而阿拉伯本地语是低变体；官方的、宗教的、广泛交际的功能按双语制的模式分布在在两种变体之中（古典语用于官方的、宗教的功能，本地语用于广泛交际的功能。”另外还使用黑体来表示“在全国占统治地位的”主要语言；用大括号和加号来表示一些包括有几种既不属于主要语言，也不属于小语言或特殊地位的语言的语言集团。

Stewart（1968）还进一步修订了他 1962 年所提出的标记系统，在语言类型的定义中加进属性，他提出 4 种属性，它们的不同组合可以标志出 7 种语言类型。这是明显地受了成分分析法的影响。

标准化，接受一套指定的正确用法的规范，并把它语码化。

独立性，该语言系统的地位是独立的，无须参照另一种语言来讨论它。

历史性，接受该语言变体为一种跨越时间正常发展的语言。

生命力，存在一个使用该语言变体为本族语的非孤立的社区。

但是这些企图用定性公式来表示语言状况的做法并未被广泛接受，其中的一个原因是应用范围不大，全世界只有 100 到 200 个国家，直接用文字来描述也无不可；更重要的原因是这种定性的归纳不够准确。Fasold（1984）在讨论这个问题时援引了化学元素的发现来说明，即使研究的范围不广，也能使我们对这个范围内的成员作出准确的排列，对原则作出不断的精细的修订，使我们对现象的了解更为深入。化学元素的研究在两条原则的基础上进行的：一条是自然主义的原则，一条是预测的原则。

自然主义的原则意味着：研究的对象是一种可观察的现象，科学家的任务是了解它是如何运作的。研究化学元素就是使用这样的方法，不少元素都可在地球上找到。但是在研究社会组织时，自然主义方法就没有那么有效。类型学和定性公式都要应用到某一个单位，而所选择的单位往往是民族。民族可以是一个自然形成的社会单位，也可以不是。因为民族的说法比较新，而且有些事实使我们怀疑我们能否应用自然主义的原则，以民族作为单位。在印度就有亚民族的政治单位——邦，可以应用 Ferguson 的公式。在某些国家，尤其是一些岛国里，地理上的差别可以导致完全不同的公式。另外的一个问题是一些种族集团并非定居在一个国家的范围里面的。Lapps 族散居在瑞典、挪威、芬兰和俄罗斯，都说 Lippish，他们在这

些国家里都是少数民族，按照 Ferguson 的说法，在那一个国家的公式里都反映不出 Lippish 的地位。自然主义可以把公式的系统同时应用到整个国家、国家的各个部分、跨国家的地区（如印度次大陆、斯堪的纳维亚和利比利亚半岛）。

自然主义对我们研究与功能有关的领域也有启发，例如官方功能通常都是指经国家的宪法和法律所规定为官方使用的语言，哪怕在实际上并没有起到这样的功能。但是在印度在1967 年法律规定以前，英语早就是官方语言；而爱尔兰虽然已经爱尔兰宪法规定为官方语言，却还没能起到这样的作用。教育语言也有这样的情况，学校当局可以规定一种语言为教育语言，但是实际上教师可能在课堂上用另一种语言。

预测的原则并非以 60 年代的定性公式和类型学为基础；提出这条原则的目的是为了分类和比较。如果我们不需要一条什么组织原则来进行预测，可供使用的系统是很多的。例如我们只需要一种分类系统来把所有的苹果进行分类，用大小、颜色、产地或其他的什么范畴都可以。可是如果分类系统必须和植物学的组织原则相一致，那么可接受的分类范畴就会大大减少。同样的，现存的公式和类型学只是按民族来进行语言分类，并且用它来比较各个民族的语言。我们也可以提出很多别的系统来进行语言分类；只要对所有国家都应用同一系统，也可以进行比较。但是只有使用组织原则来进行预测时，我们才有可能谈论正确的或错误的系统，并有可能找到王确的系统。

表 8.8　　功能和所需要的社会属性

功能	所需要的社会属性
官方	(1) 足够的标准化
	(2) 受教育的公民中的干部都知道
民族	(1) 对相当部分的人口都是民族认同的象征
	(2) 在日常生活中广泛使用
	(3) 在国家内被广泛地口头使用
	(4) 在国家内并没有其他主要的民族语言可以代替
	(5) 作为真实性象征而被接受
	(6) 与国家过去的光荣历史有联系
群体	(1) 所有成员在日常会话中使用
	(2) 团结和分裂的机制
教育（规定的水平）	(1) 学生都认识
	(2) 足够的教学资源
	(3) 足够的标准化
广泛交际	(1) 可作为第二语言而“学习”
国际	(1) 列在潜在的国际语言的单子上
学校科目	(1) 标准化程度等于或超过学习者的语言
宗教	(1) 古典语

还有一条似乎并没有指导化学元素研究的原则，也值得一提：连续统的原则（the continu-

um principle)。定性公式的另一个问题是，所建立的范畴过于呆板。例如“标准化”或“标准语”的概念。一种语言只要是标准化了的，它就是标准语，否则就是本地语。但是标准化有程度之分：Kloss 提出有 5 种不同程度的标准化。

1. 成熟标准语。所有现代知识都可以在大学里用它来传授。

2. 小群体标准语。虽然已经建立了一些规范，但是使用它的言语社区很小，现代文化难以用它来传授。

3. 年青标准语。在最近才标准化，可用在小学教育里，但还未用在较高级的研究中。

4. 非标准化的字母化的语言。并没有词典和语法，但已有文字。

5. 前文字的语言。很少或从未用诸于文字。

各个国家使用不同的语言来完成语言功能的程度是很不相同的。

根据上述自然主义、预测性和连续统这三条原则，并接纳了 Stewart 关于社会属性的观点，Fasold 提出一种新的公式系统，它考虑到两个方面：一是共同的（或接近共同的）语言功能；二是完成该功能的社会语言属性。通过比较一种功能所要求的社会语言属性和语言所赋予的完成该功能的属性，我们就有可能预测该语言完成该功能的成功率。

所谓作出预测就是把语言的属性和功能要求的属性加以比较，如果作为某一用途的语言缺乏所要求的属性，我们就预测它不能起这样的作用。以巴拉圭的 Guarani 语为例：

社会政治组织：巴拉圭
功能：民族语言
语言：Guarani

要求的属性	具有的属性
民族认同的象征	—
在日常生活中广泛使用	+
在国家内被广泛地口头使用	+
在国家内并没有其他主要的民族语言可以代替	+
作为真实性象征而被接受	+
与国家过去的光荣历史有联系	(+)

Guarani 既然具有所要求的所有属性，我们就可预测它能满足一种民族语言的要求。

8.4.2. 相关性研究

相关性研究是广义社会语言学经常采用的另一种研究手段。和上面的定性研究方法不同，相关性研究采用的是定量方法。它把语言范畴和社会范畴看成是两个密切联系但又独立的系统。语言范畴指的是用来传递关于个人物质环境的信息的言语手段，而社会范畴则是这个物质环境的组成部分。相关性研究的出发点是：把这两套独立测量的变量作相关分析，我们就可以看到社会结构和语言结构的系统变化。社会范畴是按其独立于交际过程的社会特征来测量的，它包括社会经济地位（一般的收入、教育、职业所决定），出生地点，所属的群体，年龄，在评估测试中的态度，等等。在进行相关分析时，社会范畴是自变量（independent

variables)，语言范畴是依变量（dependent variables）。例如在下面要介绍到 Labov 的著名的调查里，/r/ 是一个标志社会地位的变量："如果我们把纽约市的两个亚群体的说话人在一个社会阶层的量表上排列，他们在使用/r/时的差异也是按同一次序排列的。"（1972）这也就是说，社会地位越高，使用/r/的情况就越多。

在美国，在 50 年代初期就开始有人对语言和社会特征的相互关系作系统的观察，主要是研究美国最低层的黑人的英语特征。但是 Labov 认为这些调查不够严密，首先是样本的挑选不够系统，其次是要考察的语言特征不够清楚。

Labov 的调查引进了新的方法，首次对言语行为作出准确的，以实际数据为基础的分析，对语言的变异提出了新的解释。他的研究的特点是：按照定义清楚的社会范畴来选择被调查人（见 8.2.2），区别特定语境的语体，了解语言变量和语言以外的参数之间的相互关系。

Labov 的样本包括了纽约市的各个民族群体（纽约人、意大利人、犹太人、黑人）和不同的社会阶层：中上层（UMC）、中下层（LMC）、上层工人阶级（UWC）和下层工人阶级（LWC）。他就下列几个方面一共访问了 122 个对象：

1. 详细的社会背景资料：关于被访者的收入、职业、出生地、年龄、宗教、母语、个人地位等等方面的资料。

2. 词汇：日常生活中常用物体的词语，目的是记录词汇使用的差异。

3. 社会风俗：被访者在什么环境里长大，他参与哪些一般性的社会活动。

4. 句法和语义：言语形式的特殊性的资料；设法让被访者讲一个故事，并对此作语法分析。

5. 发音：朗读、读词汇表、读最小的语音对立的对子（如英语的 god/guard）。

6. 语言规范：在两种不同的语言变体中，让被访者按照他们的意见来决定哪一种是正确的。

7. 言语反应测试：使用反应测试在纽约人中进行言语采样；让他们使用自我评估测试来评估自己所说的话。

8. 语言态度：对纽约以外的语言变体进行评估。

Labov 对每一次的访问都进行了录音。如果其他家庭成员在访问中介入，也对他们（特别是小孩）进行访问。由于技术方面的原因，只能对 33 个被访者进行简短的访问，但是在访问中让他们回答了最主要的问题。

Labov 选定了 5 种特定环境中使用的语体（见 8.3.2.3），最后两种（词汇表和最小对立体）其实可以归为一类：词汇表。在调查中，随便的语体的性质较难界定，而且要使用一些技巧才能诱发出来。这 5 种语体是一个从非正式到正式的连续统。

Labov 根据 4 条标准来选定语言变量：(1) 使用频率很高；(2) 不会受到故意压制的影响；(3) 能够结合到更大的结构里；(4) 能够在线形的量表上量化。结果发现有 5 个变量可以满足上述要求：(oh)，(eh)，(r)，(th)，(dh)。把这些变量加以量化，就能表示社会、民族和语体的差异，所以我们把这种研究称为相关性研究。如果我们把说话人属于不同的阶层理解为他们使用了不同的社会规范，把他们使用语体上的差异作为不同程度上的言语监察的反映，那么语言变量是在两个维度上产生差异：

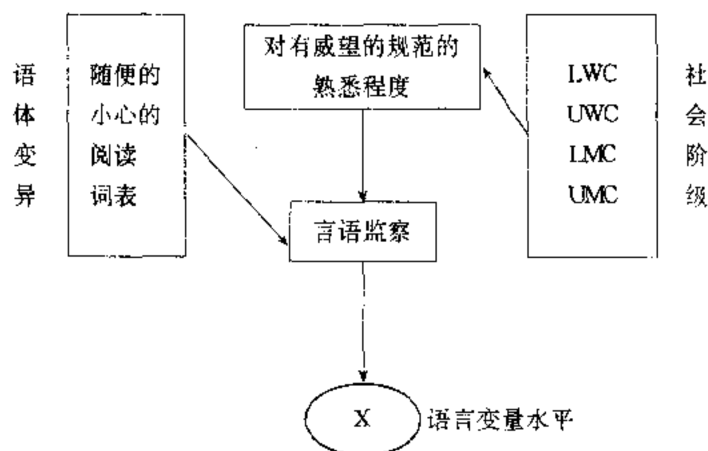


图 8.2　介入语言变量的不同水平

这样，语言变量在两个维度上区别出社会距离和语体功能。例如英语的辅音/r/处于在元音后的位置，如 beer，bare，car，four，moor，moored，board，fire，flower，flowered，有不发音 (r-0) 和发音 (r-1) 两种变体。Labov 认为它是社会和语体分层中的一个很敏感的标记。任何两组纽约人在/r/发音上的不同的排列次序和他们在社会阶层的量表上排列的次序是一致的。

Labov 在纽约第五大街的 3 间代表了高、中、低消费水平的百货公司 Saks，Macy's 和 S. Klein 对售货员采取了 8.3.2.3 所介绍的方法进行快速访问，然后把结果分为 3 类：

(1) 只用 (r-1)，没有 (r-0)

(2) 有些 (r-1)，起码有一个 (r-1) 和一个 (r-0)

(3) 没有 (r-1)，只用 (r-0)

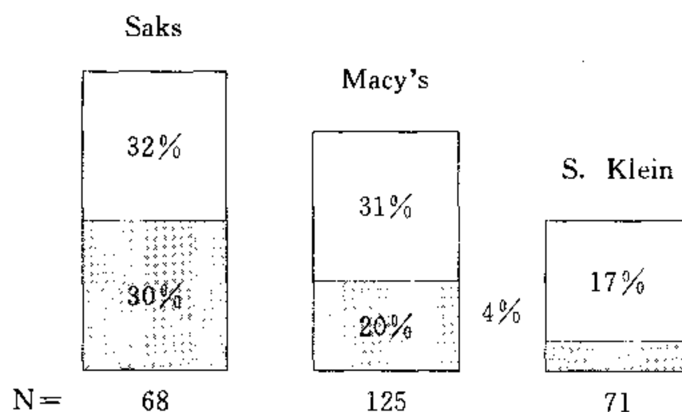


图 8.3　三类用语比例

在图中有阴影区为 (1)，即只有 (r-1)；没有阴影区为 (2)，即有些 (r-1)。图中没有表示 (3)。由此可见，在高消费水平的 Saks 里，62%的售货员全部或部分使用 (r-1)；在中等消费水平的 Macy's 里，有 51%；在低消费水平的 S. Klein 里，只有 21%。如果我们只看全部使用 (r-1)，则差别更大，30∶20∶4。

Labov 还进一步调查了纽约市各阶层在不同语体中使用 /r/ 的情况，获得更有意义的发现：

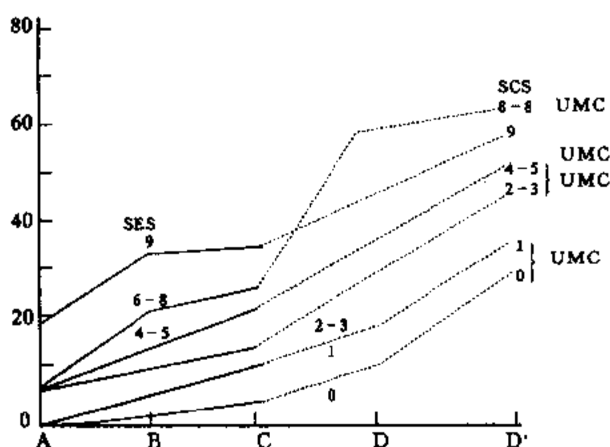


图 8.4　约约市各阶层使用/r/的情况

从图中可见，不同的社会经济阶层在不同的语体中使用 /r/ 的情况是不同的：社会经济地位越高的说话人使用 (r-1) 的越多；越是小心的语体，使用(r-1)的也越多。还有一个有趣的现象，中下层(LMC)阶级在念词表时有个陡坡，上升得最高，Labov 称之为过度正确(hypercorrection)，而且认为它是语言变化的可靠指数。这是因为在第二次大战前，纽约的地方话是 (r-0)，(r-1) 在战后成为一种有声望的形式。美国的中下层在战后的地位有所上升，急于与这种形式建立联系，因此使用 (r-1) 时比中上层还要多。这是这个阶层有意识那样做的，Labov 称为这是意识水平以上的压力。

表 8.9　　听懂的百分比和会话特征的百分比的相关

会话编号	听懂的百分比	英语同源词的百分比
9	79.7	93.9
11	79.1	95.5
17	75.5	94.3
12	74.9	97.8
16	68.9	87.3
13	65.5	93.5
15	62.5	92.0
7	60.2	95.3
9	59.1	87.7
10	58.8	93.6
5	56.0	85.9
4	55.5	91.0
3	53.4	84.6
8	50.4	90.5
6	43.6	85.5
2	34.1	90.6
1	32.5	89.5

Labov 所使用的分析方法实际上是对语言变异的分析。从说话人的角度看，视其社会经济阶层、性别、年龄、教育水平、职业、地区、民族之不同而出现语言使用的差异；就语言本身来说，有语音、语义、句法、语体的差异。甚至同一个人在不同的场合也会使用不同的语

言变体，这就是语码的转换。近二三十年来，这些方面的调查和研究很多，在这里不能一一介绍。但是Labov是从语音调查开始的，而且他所提出的说明力原则(the principle of accountability)有很大的方法论的意义。这个原则认为分析家不能仅从资料中提取那些符合自己的观点的语言变异情况，而忽略那些不符合自己观点的情况："就所考察的那部分言语而言，某一种变异的所有出现情况都必须记录下来，而且只要我们能够说出属于这种变异集合的那些变量在相关的环境里没有出现的情况，也要记录下来。"换句话说，对言语社会中语言形式或规则的报告必须包括观察所取样的总体的话语的说明，包括这个形式在期望的环境中的发生率。

虽然Labov是在观察社会经济阶层的变化和语言变化之间的相互关系，但他所使用的方法并非统计学上的相关分析法。Flint (1979) 调查了南太平洋澳属Norfolk岛居民在本地语接触中的相互了解的情况。他让一个说澳大利亚英语的人去听17篇以英语民族混合语为基础的会话，要求他逐句报告其意义。报告人是一位语言学家，对民族混合语很熟悉，但不懂得这种语言。在下表中列出他听懂会话的百分比，并列出会话中其他特征，如该会话中与英语同源、意义也接近的词的百分比。

用统计学方法求出两组数据的相关系数 (r) 为0.544，[①] 说明它们具有中等程度的相关。相关系数的乘方 (r^2) 是表示所测量的两组数据共同享有的方差的统计量，在这里意味着在Norfolk岛居民的会话里，英语词的百分比可以解释互相了解程度的29.6%的方差。

8.4.3. 交互作用分析

交互作用分析法的出发点是社会中的交际活动不是一些毫无区别的话语字符串，而是各种各样语言事件。它们有不同的范围，表示不同的行为规范（包括使用不同的语体）。因此对这些交际活动的描写必须列举一个社区里对其语言事件的分类，列举作为这些事件的开始和结束的标记的性质，列举区分不同事件的特征。

每一个语言事件都有其开始和终结。一个事件的结束后，可向另一个事件转移，或导致交际的中止。电话中的会话以呼叫开始，以挂上电话而结束。一个故事的开始是"从前……"，结束是"从此以后……"。祈祷的开始是"让我们祈祷……"，结束是"阿门。"正式讲话的开始是"各位……"，结束是"谢谢！"事件的界限有时可用（或辅以）脸部表情、音调的高低、一个事件和另一个事件之间的身体移动或一小段沉默来表示。最明显的界限是语码的转换，或刚好碰上话题的或交谈者的改变。

交际有其规范，违反了就会引起他人的反感或不安。在音乐会上演奏未结束就鼓掌，说明对事件的结束有不同的看法；在奏序曲时私语，表示说话人并不认为音乐已经开始。

8.4.3.1. 交际的组成部分

我们可以首先分析一个交际事件的组成部分：

1. 场景 (Scene)，就是背景，它包括类型 (genre)，话题 (topic)，目的或功能 (purpose

① 他使用了统计学中的积差相关系数，具体的公式请参看11.5.2。

/ function)。场景是事件的外部环境，只有它可直接观察。有时观察者对不属于自己文化的背景细节往往会忽略，例如在日本，椅子的相对高度对了解事件的意义至关重要；而在英国的课室里，把课桌排成直线和圆形意味着正式程度的不同。

在场景中的事件类型很不同，如讲故事、演讲、开玩笑、问候、讨论问题，等等。在一天的什么时间，一个星期的哪一天，一年的哪个季节说话都会影响语言形式的选择，某些类型只适宜于在某些时间使用，Navajo 人只在冬天讨论动物冬眠，而犹太正教徒在星期天不能讨论世俗的事务。时间和地点都会影响问候的含义，中国人的“吃了饭没有?”也是一种问候的方式，但只能在某些时间说。

场景的有些部分不能完全地直接观察到，例如一些自然现象（星星和月亮，打雷和闪电）对不同的民族有不同的含义，各个民族的风俗习惯和信念也很不相同。

关于场景需要回答下列的问题：

· 它是什么类型的交际事件？这个交际事件是关于什么的？它为什么会发生？它在什么时间和地点发生？它的背景像什么？

· 人们是怎样组织空间（如排列、圆圈、围着桌子、在房间的中间、围着圆周）以达到不同的目的？人们有些什么地理空间概念、认识和信念？东南西北那样的方向对他们有什么意义？不同的方向和地点（如天上、埋头向西走、主人在吃饭时对着门口坐）对他们有什么意义？

· 人们对时间有些什么信念和价值观？和时间相联系的有些什么特定的行为和禁忌（如夏天不要唱歌以防蛇咬，太阳下山后不要讲故事)？

· 什么东西是神圣的？什么是世俗的？有些什么信念、行为、禁忌是和自然现象相联系的？

· 人们怎样度假和庆祝节日？其目的（如政治、宗教、季节、教育）是什么？哪些行为被看成是“工作”？哪些是“娱乐”？

· 哪些穿着是“典型的”？在特殊的场合穿些什么？

参加仪式有些什么外部的标志（如穿着、在皮肤上做标记)？

2. 基调 (Key)。基调说明完成一件行为所使用的调子、方式和情绪。基调往往可以对照，如：取笑和认真、真诚和讽刺、友善和敌意、同情和威胁、马虎和苦干，等等。对类型来说，基调往往是多余的，如开玩笑是可笑的，慰问是同情的；但也不一定，如用讽刺的基调来开玩笑，带有恐吓性的慰问。某一种基调可以首先和某一种语言使用功能、某一种交谈者的身份关系、某一种消息的形式和内容相联系。

基调之所以重要是因为它是一个独立于其他交际事件的变量，当交际事件的各个组成部分发生冲突时，基调可以压倒其他成分，例如用讽刺的口吻来恭维别人，讽刺的基调就压倒消息的形式和内容，传递了另一种交谈者的关系。

传递基调的方式是多样的，可以通过语言或变体的选择，可以通过非言语的信号（如眨眼和身势），可以通过伴语言特征（如送气的程度)，也可以综合使用这些手段。

基调有文化内涵，必须按照本地人的理解来解释。因为它有一种压倒其他成分的作用，在分析交际事件时，必须先加注意。

3. 参加者 (Participants)。参加者的年龄、性别、民族、社会地位和其他有关的因素，他们的相互关系是交际事件的重要组成部分，所要回答的问题是："谁在参加交际事件的活动?"我们所要描写的不仅那些看得到的特征，还要提供他在家庭或其他社会组织中的地位和身份的背景知识。参加者包括说话人和听话人，他们在言语活动中保持着一种社会身份的关系，如父/子、夫/妻、师/生、友/友、上级/下级、牧师/教友、顾客/售货员、乘客/售票员，等等。这种关系并不稳定，师/生关系很容易改变为友/友的关系，要表示这种关系的改变就要改变说话的方式。外语学习者最初只学到一种表达方式，不管是对小孩还是成人、售货员还是教授、朋友还是官员，都用这种方式，常会闹笑话。

下面是关于参加者的一些需要回答的问题：

· 在"家庭"里有谁？谁是住在一间房子里面的？在家里谁最有权威，谁次之？每个家庭成员有些什么权利和责任？家庭在更大的社会组织里有些什么作用和义务？

· 生命中各个阶段、时期、过渡是按什么范畴来定义的？在生命周期的不同阶段里对个人有些什么态度、期望和行为？哪一个阶段最有价值？哪一个阶段最"困难"?

· 谁对谁有权力？一个人能够强加于别人的程度有多大？用什么手段？

· 社会控制的手段是否视生命周期的不同阶段，所属的社会范畴的不同而有所变化？或按照环境或犯法行为的不同而有所变化？

谁在群体里可以具有什么身份？他们是怎样获得这些身份的？有哪些身份具有特别善意或恶意的特征？

下面是关于参加者和语言、文化的关系的一些需要回答的问题：

· 语言怎样和生命周期相联系？语言使用在定义社会标志和身份中是否占重要位置？

· 人们在不同的身份关系中怎样称呼对方？怎样表示尊重？怎样表示侮辱？

· 谁可以不赞成谁的意见？在什么情况下可以这样做？

· "很会说话"的特征怎样和年龄、性别及其他社会因素相联系？说话能力、识字、写作能力怎样和一个人在社会中的成就相联系？

· 有哪些身份、态度和性格特征和特定的说话方式相联系？

· 谁可以对谁说话？在什么时候？在什么地方？谈论什么内容？

语言在社会控制中有些什么作用？使用什么变体？在多语制社会里，使用第一语言和使用第二语言有些什么不同的作用？

关于谁能参加交际事件的看法也有文化内涵，而且不限于人类。西方人往往认为他们可以和自己的宠物交谈。

3. 消息形式 (Message Form)。在研究各种社会、文化和环境对交际行为的制约时，言语和非言语的语码对交际事件中的消息形式、消息内容和活动次序有重要影响，每一种语码都是通过有声的或无声的信道来传递的。言语和非言语、有声和无声形成了四向的关系，可表示为下图：

表 8.10 语码和信道的四向关系

		信道：有声	信道：无声
语码	言语	口语	书面语 (聋哑人) 手势语 口哨/鼓点语 摩尔斯电码
	非言语	伴随语言和韵律特征 笑声	沉默 身势语 近体语 眼睛接触 图画与卡通

对言语性语码的描写通常限于口语和书面语，但是其他言语交际方式也颇为广泛，有些地方使用口哨和鼓点来进行交际，在船只之间也可用电报码和旗语。另一种言语/无声的交际方式是各个社会都有聋哑人使用的手势语，在手势语中还伴以一些非言语的方式，如面部表情、眼睛接触，等等。

在语言学中，沉默常被人忽略（除了作为话语的分界）。Saville-Troike（1982）指出，沉默有两种：一种是有意义而无命题内容的沉默，包括在轮流说话时的沉默和停顿。这可以说是沉默的韵律作用。这些无命题内容的沉默可以是有意的，可以是无意的，具有各种不同的意义。它所传递的意义往往带有情感，强调的是意义的外延，而不是其内涵。所以这些意义是符号性的，往往是约定俗成的。另一种是完全依赖邻近的声音赋予意义，本身具有施为作用（illocutionary force），可称为沉默的交际行为。这些行为是有命题内容的，它们包括手势；但也可只有沉默，而无任何可见的提示。甚至在不可能有可见信号的电话交谈里，对哪些要求有言语反应的祝贺、提问、请求而作出沉默的反应，都有其命题内容。正如人们可以念念有词，而不说什么话，我们也可以不念词而说话。沉默也是一种"言语"行为，起到其他言语行为所能起的作用。要研究沉默的交际行为，可观察沉默是怎样具有语法和词汇意义以取代话语中的不同成分，例如在课堂里，教师不问"这是什么?"而说"这是________?"和新朋友相交，用"你的称呼是________?"来代替"怎样称呼你?"当话题比较微妙，或要说的话是禁忌的，或说话时过于激动，说话人"说不出话来"，话语往往以沉默而结束。用沉默来表示不同意或反对，是一种较客气的表现。在一些宗教仪式里，往往需要对整个交际事件都保持沉默。

从方法论的角度看，对不熟悉的文化的描写往往忽略了沉默，因为有些语言学家对沉默持否定态度，认为它是缺乏任何特征。Saville-Troike 认为，我们必须对沉默具有一种"元意识"（meta-awareness），注意各种可能出现的沉默，对它们的解释要特别小心。

和沉默类似的另一种现象是"后信道"(backchannel)信号，它包括听话人在交往中的各种反应，例如在英语中的非言语的声音：mm，hm；言语的声音：yeah，I see；无声的点头和身体移动。这些信号有不同的示意：消极的承认、积极的鼓励、希望转换话题或轮到对方说话。笑声也常为人忽略，但是它和事件类型、话题、基调和其他成分相联系，也有不同的型式。

在描写像身势语和面部表情这样的非言语/无声的行为时，应注意(a)身体的哪一个部分是在移动或处于显著位置；(b)移动的方向以及它和非显著状态有什么不同；(c)移动的范围。

4. 消息内容(Message Content)。消息形式和消息内容是互相联系的，紧密不可分的。消息内容指所谈论的是什么交际内容，所传递的是什么意思。例如：

甲：听说有4个乡亲明天会到我们家来，但是我们住得太挤，没有地方安排。
(隐含的请求是“你能帮个忙吗?”)
乙：[沉默，面部没有任何明显的表情](隐含着拒绝：“我也无能为力。”)

甲的话和乙的沉默都是消息形式，而隐含的意思则是消息内容。在面对面的交际里，我们不但从言语的和非言语的消息形式和消息内容中获取意思，而且也根据超语言环境和参加者带进交际事件的信息和期望来推断意思。这些因素都是同时处理的，我们很难再把它们细分来分析。为了观察非言语因素在交际中的作用，Saville-Troike主张研究互相不懂对方言语的交际事件，下面是一个刚到美国的中国小孩(P1)和一个不懂的汉语的美国托儿所教师(P2)的交际：

(a) P1：我的鞋带断了。
P2：Here you go.(好啦)[她把鞋带打个结]
(b) [P1拿着一个破了的气球]
P1：看，看。我这没了。看，看。
P2：Oh, it popped, didn't it? All gone.(哦，爆了，是吗？都没了。)：
(c) [P1在看着洗涤槽里的水]
P1：怎么这个水都不会流啊？
P2：It fills up, uh huh. It doesn't drain out very fast, does it?
(哎哟，都满啦。水流得不很快，是吗?)

从上面几个例子里可看出，因为有具体的对象和一个双方注意力都能集中的特殊条件，大家很容易相互理解。P2处于这些外部环境，加以接触小孩的经验使她能够期望小孩在该环境里会说些什么，故对P1话作出了合适的反应。在下面一个语义不很连贯的例子里还可看出期望的重要性，老师把一幅狗的图画给小孩看，她期望他们会讨论他们经验中的狗，因此她就以为P1的话是关于他自己的狗，他的手势是表示狗的太小。但是P1谈的却是恐龙，而平面的摆手是说明地理结构的。教师在这里推断不出消息内容，因为它超出了她关于托儿所小孩会谈些什么的“期望结构”：

P1：恐龙好久，好久。恐龙现在已经变成煤矿啦。
P2：Do you have a dog with you?(你有一条狗吗?)
P1：很深噢。一拨一拨一拨的。本来在这边，地势这样起来。搞到这边。
[P1平面摆手，以说明地势]
P1：恐龙在这边。

P2：Oh，Growing big.（噢，长大了）

这几个例子说明，成功的交际依赖于正确传递和理解消息内容，而成功的交际实际上有程度之分。就共享话题和了解对方意图而言，头3个是成功例子的，而后一个是失败的例子。虽然这些例子说明缺乏共同的语码，在某些可高度预测的环境里，消息内容仍然可以传递，但也有不少例子显示，由于没有共享语际知识和期望，即算说同一种语言，仍会出现人们误解消息内容的情况。

5. 活动次序（Action Sequence）。活动次序指的是一个言语事件中交际活动的先后次序。在这个次序里，一个参加者的活动后面跟着另一个活动，前一个活动为后一个活动建立环境，而后一个活动确认前一个活动的意义。在一些诸如祝贺、赞扬、吊唁的仪式里，活动次序是比较严格的，但在会话里，则比较自由。

在描写活动次序时，我们通常是按照功能来表示交际活动，附以典型的消息形式和消息内容，例如Tsuda（1984）在观察了日本23个上门推销商品的案例后，找出一个典型的型式：

P1（推销员）：问候。

对不起。（原为日文，这里只给出汉语意思。下同）

P2（主妇）：确认。

是。

P1：表明身份。

对不起。我是J公司的。是的，J公司。

P2：询问来访目的。

你找我干什么？

P1：告诉目的。

太太，您知道那电视广告吗？那个既能缝厚的，又能缝薄的……。

P2：表示感/不感兴趣。

哦，缝纫机。我们家里已经有一部了。

在这个层面上的概括不但可以显示规则的型式，而且还可以让我们进行跨文化的比较。就以上门推销为例，日本和美国基本上是一样的，但在形式和内容上也有一些显著的差异，推销员在美国通常是先介绍自己的名字，不像日本的推销员那样先介绍自己的公司。

6. 交往规则（Rules for Interaction）。这一部分对在交际事件中言语的使用规则作出解释。这里所说的规则指的是一些描写性的陈述，以说明人们怎样按照他们在社区中共享的价值观行事和他们认为应该怎样做。这些规则也可进一步描写一些典型的行为，但不一定非要这样做。这些规则可以体现为格言、成语或准则，但是也可以无意识地执行。对交往规则的认识往往是因为它们被人违反了，或引起了一些“不礼貌”或“古怪”的感觉。例如会话中的轮流说话的规则有下列三条：

(a) 正在说话的人对谁说话，就轮到谁说话。

(b) 谁先说话，就轮到谁说话。

(c) 如果正在说话的人在别人说话前自己接着说话，就论到他说话。

这 3 条规则是有先后次序的，(a) 要优先于 (b)，而 (b) 又优先于 (c)。如果甲正在对乙说话，就该乙说话，丙不能执行 (b)，抢先说话。

Wolfson (1980, 1981) 对美国英语中说恭维话作过有趣的研究，她认为说恭维话具有一种加强说话人和听话人的团结的社会功能，而且有规律可寻。根据她对约 700 个恭维话的语料的调查，有 23%用 nice，有 19%用 good，用 beautiful，pretty 和 great 的各占 5%强，所以这 5 个形容词约占 2/3。就句型而言，最常用的只有 3 种：

(a) NP [is] / [looks] (really) ADJ

(b) I (really) [like] / [love] NP

(c) PRO is (really) (a) ADJ NP

(a) 占 53.6%，(b) 占 16.1%，(c) 占 14.9%，三者合共占 85% 。因此她不得不下这样的结论：美国英语的恭维话的显著特征是几乎完全缺乏创造性。

7. 解释规范 (Norms of Interpretation)。解释规范提供所有其他了解交际事件所需的关于言语社区及其文化的信息。把它称为解释规范是因为它们是言语社区成员共享的标准，例如一个马里的 Bambara 村民在集会上辩护自己的观点必须使用直接引语（简明扼要），而表示反对就必须用间接引语（谜语和寓言）。

8.4.3.2. 交际事件的分析

分析交际事件的部分任务是了解在一个特定的言语社区里，有哪些组成部分是起决定作用的。在第一个阶段，我们使用一个框架来收集数据，找出哪些本地人认为是有意义的差别。最简单的框架多少有点像结构语言学那样，使用“最小对立体”来进行比较，例如在了解问候的言语事件时，调查者可以注意和记录几种不同的问候方式，然后从消息形式、内容、参加者、基调和场景等方面加以比较，然后再向参与交际活动的人访问，看他们是否也认为各种问候方式在含义上有所不同。调查者还可以让整个框架作为常量，只作最小的改变，以了解它对交际行为有些什么影响。例如：如果一个参加者比别人的年龄都要大，问候方式会怎样？如果一个参加者是男性，另一个是女性，又会怎样？如果一个妇女带着面纱，会不会有所不同？如果不是在早上，而是在晚上，或不是在房子里面，而是在街上，会不会又有所不同？

另一种更为复杂的发现程序是要求资料提供人扮演角色，让他设想自己处于某一特定的环境里活动，然后调查者观察其行为有些什么不同。扮演角色所产生的往往是“理想化”的或型式化的行为，必须在自然观察中加以验证，才能有效。Laughlin (1980) 在墨西哥 Chiapas 的一个马亚人的社区收集关于交际环境的数据，发现村落里喜欢闲聊一些偷情和私奔的话题，但他却难以就这样的内容向女孩子进行访问，更难以直接参与。后来他想出一条妙计，给他的资料提供人 3 个题目，让他提供场景和假想一个男子和他的女友的对话，于是他就收集了“想像中的勾引女孩子”，“想像中的已婚男子勾引寡妇”和“想像中的酒徒勾引女孩”的资料。

根据框架来进行分析仅是第一步，合适的分析必须跳出静态的框架，进一步考虑使用交互式模型的框架，可称为动态的“图式”或上文提到的“期望结构”。这种方法认为，人们并非作为天真的、空白的容器去接近世界，把什么东西都作为独立的和客观的物体来接受，他们都是经验丰富和老于世故的感受能手，存储了很多知识，他们根据自己先前的经验和对事物之间的相互关系的理解来对待世界上的各种事物。这些先前的经验表现为对客观世界的各种期望；客观世界在大多数的情况下也是有组织的，能够确认这些期望。这样我们就不必要对每件事都从头开始观察。

这样，我们必须了解说话人所使用的是什么框架，他们是怎样把期望和语言活动联系起来的，他们使用什么图式和交互过程去对待共享的文化经验，才能达到解释交际能力的目的。但是用什么方法去收集和分析这些信息却是一件富有挑战性的工作。

有的人（如 Chafe 1980，Tannen 1981）采取的办法是把一个记录片放给 10 个不同国家的调查对象看，然后要求他们描写片中的内容，再根据他们在复述中组织事物的方式来推断期望结构是怎样受文化影响的。各种交际场合的记录片和照片都可以用来向参加者提取他们对这些场合的解释。另一个方法是让所调查的群体中的一个成员来执镜拍摄，因为拍摄者总要选择镜头和取景，这样就可以了解到他的焦点所在。

Saville-Troike 组织了美国乔治敦大学和伊利诺斯大学的一些外国留学生来根据上述的框架描述他们本国的一些交际事件，下面是几个案例：

1. 操 Bambara 语者在马里的一个传统的村落集会上的交际事件

话题：怎样防止动物侵犯庄园

功能/目的：作出管理村落生活的决策

场景：如果是正午烈日，在树底下举行

如果是下午或黄昏，在村公共集会处举行

基调：严肃

参加者：村里所有男性居民

P1：酋长

P2：传令官

P3：活跃居民（45 岁以上）

P4：半活跃居民（21—45 岁）

P5：不活跃居民（14—20 岁）

消息形式：Bambara 口语

P2 大声说话；其他人用温和的声调

活动次序：

P1 念议事日程

P2 向大众宣布议事日程

P3（其中一个）请求发言

P2 把要求传递给 P1

P1 同意或拒绝请求

P2 把同意或拒绝请求传递给 P3

P3 提出意见（如果 P1 同意）

P2 把意见传递给 P1 和大众

（当活跃居民（P3）轮流发言时，重复 3—8 的活动程序）

P1 总结讨论情况，提出建议

P2 把总结和建议传递给大众

交往规则：

只能有一个活跃分子（45 岁以上）说话

可以向半活跃分子（21—45 岁）征求意见，但他们不能主动说话

每个说话人必须向首长请求允许发言

首长和其他参加者不能直接交谈；传令官把首长的话传递给大众，或把一个人的发言传递给首长和大众

为了影响别人或显示自己的重要性，活跃居民必须轮流发言

解释规范：

直接引语（简明扼要）意味着说话人在辩护一个观点

间接引语（谜语和寓言）意味着说话人反对一个观点

集会中的人是严肃的

传令官不一定是严肃的

2. 在象牙海岸的操 Abbey 语者问候的交际事件，这个例子说明性别和年龄的不同会产生一些变异，主要是体现为活动次序的不同。问候的背景对内容和次序也会产生差异，但在下例里，这个成分是一个常量：

功能/目的：在访问的开始，重新确认参加者的友好关系

背景：私人住宅

基调：友善

参加者：

P1：住宅主人

P2：来访客人

变量条件

A. P1 和 P2 都是男性成年人，或 P1 为男性，P2 为女性

B. P1 为女性，P2 为男性

C. P1 为小孩，P2 为成年人

D. 同时来了几个客人

活动次序

条件 A

第一步——“问候和回应”

P2 问候

P1 接受问候

P1 找椅子给 P2（如果找不到，就会使问候的次序有一个长时期的停顿）

第二步——“请坐”

P1 给 P2 一张椅子

P2 致意

第三步——“打听消息”

P1 和 P2 坐下

P1 向 P2 打听消息

P2 给予标准的公式化的回应

条件 B

第一步和第二步和条件 A 相同

P2 匆忙地找寻最接近她的男子去完成问候次序

如果找不到，她就违反规则，道歉，通过“打听消息”来完成问候次序

条件 C

如果 P1 是一个年纪小的小孩，不需要问候

P2 要 P1 去叫其父母

如果 P1 是一个年纪大的小孩，可在找父母前完成第一步和第二步

条件 D

客人中最年青的成年人负责传递消息

在第三步，P1 直接和这一群人中指定传递消息的人交谈；这个人必须和其他人商量后再作回答

交往规则：

十岁以上的小孩都有“权”接受问候

朋友之间的问候次序没有严格要求，但是一个总是首先问候的妇女可能会被人看不起。

解释规范：

如果第一步和第二步被略去，或次序有所改变，P1 和 P2 之间的关系就不大正常

“打听消息”是问候的一部分，并非访问的目的

谈论消息以后，P2 才会说出访问的实际原因（开始另一个言语事件）

3. 日本的求婚，一个只包括一句话的交际事件

功能/目的

宣布结婚意图

建立或发展合适的身份关系

基调

严肃

参加者

P1——男性；年青成年人

P2——女性；年青成年人

消息形式

言语——口头日语；沉默

非言语——身势语；眼睛注视

消息内容和次序

P1　拿着 P2 的手（非必要的）
　　看着 P2
　　说“请嫁给我”
P2　低着头
　　沉默

交往规则

男子必须向女子求婚

在感情处于高潮时应保持沉默

女子的头应低垂，她眼睛注视的方向应低于男子的注视方向

解释规范

男子是一家之主，应主动求婚。这个习惯来自早期日本神话，当妇女之神和男子之神结婚时，是妇女之神首先提出，可是他们婚后只能生育像虫子那样的邪恶之物，所以他们再结一次婚，这次由男子之神提出，婚姻成功，生育了一个叫做“日本”的国家。习惯保存至今，大家认为这个规矩不能被破坏。

人们还相信，当一件事用词语来表达（口头或书面）时，其真正的本质就丧失。在父母去世，儿子通过大学入学考试，或看见一些特别美丽的东西时，都应保持沉默。有一首著名的诗歌，开始是“哦，松岛……”，诗人被海岛的美丽所吸引，无法继续。该诗为经典之作。

结婚是女子生命的高潮和主要目标，求婚如此重要，最合适的反应是沉默。

低首和目光向下是谦虚的表现，而谦虚乃女子的一种珍贵的德行。这种反应是一个年青男子所期望的，它确认他要娶的正是这个女子，他们未来的生活将会是平静的，而他会是一家之主。其实他不是在向她提问，要求回答，他在宣布他要娶她的决定。

4. 在美国的中国留学生请人吃便饭的交际事件

功能/目的：

增加友谊

对别人的帮忙表示感谢

背景：

P2 在大学的办公室，下午五时

基调：

友善而随便

参加者：

P1——中国研究生，男性

P2——中国研究生，男性

P1 和 P2 来自中国的同一个城市，通过亲戚而互相认识

P2 最近回国作短期访问，代 P1 的父母给 P1 带回一些东西

消息形式：

标准汉语口语，北京话

随便的语体，在话语中插入感叹语；升调和降调

头部动作（点头、摇头）；面部表情

内容和次序：（由几步组成）

第一步　开始

P1 问候

P2　接受问候

请坐

还以问候

第二步：邀请

P1 提示他想请 P2 做件事

停顿，看 P2 的反应（面部表情）

邀请到他家吃饭

P2 拒绝邀请（惊异感，皱眉头）

P1 坚持对方接受

P2 间接接受（面部表情说明他无别的选择）

P1 一再表示邀请的诚意；商定具体时间

P2 同意该时间；表示感谢

P1 一再说明是便饭

第三步：结束

P1 确认时间

找一个离去的借口

P2 对 P1 一再感谢

临别致意

P1 临别致意

交往规则：

请客的人应坚持两到三次，但应视被请人的反应而有所控制。

在接受邀请前应拒绝两到三次：首先谦虚地拒绝，然后间接地接受

通过面部表情来表示不想接受，然后因为无他选择而接受。

解释规范：

在中国，请吃饭是一件重要社交活动，有两重功能：(a) 增加社会联系；(b) 表示感谢或表示需人帮忙。

请客的人按照他观察到被请人的面部表情和话语的措词和音调而决定坚持邀请的程度。

如果被请的人的面部表情是犹疑和冷淡，或是直接说“不”，或是找到一个很好的借口，请客的人就不要再坚持。

接受邀请的方式反映一个人的态度和自律：先是谦虚地拒绝，然后间接地、带有沉思地接受，这被认为是有礼貌的、态度良好的、周到的举动；反之就是没有礼貌的、态度不佳的。

以上的分析仅是一些个案记录，基本上是描写性的，并不说明所描写的事件就是规律性的。如果要得到规律性的结论，那就必须有较多的人对同一事件进行描述和概括，看是否一致。

8.4.4. 数据的形式化描写

从 Labov 开始，社会语言学家在语言数据的描写方面采取了形式化的手段，一个原因是 Chomsky 语言学的影响，另一个原因是社会语言学家认为语言变异和社会的关系应该上升为规则，而且只有经过形式化处理才有可能实现计算机人工智能的前景。这是社会语言学和人类语言学的差别之一。社会语言学企图用变量规则（variable rules）来反映在社会各个因素的作用下语言变化的规律。

按照 Labov 等人的看法，描写语言数据有 3 种不同类型的规则：

1. 绝对的规则（categorical rules），亦称不变的规则（invariant rules），大部分规则均属于这种类型。这些规则不会被违反，因此也很难去定义。它们对说话人来说，是“无形的”：当说话人听到违反了绝对的规则的句子，他们很难解释“这样说究竟是什么意思”，他们的反应是“你在 x 语里不能这样说。”

2. 半绝对的规则（semi-categorical rules）。这类规则是可以违反的，可解释为“也可这样说”。虽然它们的使用频率并不高（如增加表现力，赋予感情色彩，提高修辞效果），它们还算是语言的潜在表达方式，足以引起人们的注意。人们对这些说法的反应是：“他是真的这样说吗？”

3. 变量规则。第三种类型的规则是个别话语不能违背的。它们是分析家经过调查才发现的。听话人仅是潜意识的感知这些规则，并根据使用规则的情况提供关于说话人信息（如性别、受教育的情况、来自何处）。一般来说，说话人不能直接说明这些规则。

8.4.4.1 变量规则

由此可见，提出变量规则的出发点是反对把任何语言规则都说成是绝对的规则，认为语言的使用存在着有规律的变异，没有这些语言和语体的变换，就不可能有交际。Labov 虽然不赞成生活语法那种过于依赖个人的做法，但是他所提出的变量规则却是建筑在 Chomsky 1965 年的生成转换语法的模型上面的。换句话说，他认为每一个句子都有其深层结构，要经过转换的步骤才能变为表层结构。

John *run/runs* every day.

在标准英语里，有一条关于数的一致性的规则，所以要说 runs。但在英语的一些变体里，这条规则却不是绝对的，即可以说 runs，也可以说 run。不过这条规则的使用与否，也并非任意性的。相反的，这条规则的应用及其应用的频率在很大程度上是这个变量（数的一致性）的语境和言语的环境（说话人的地区和社会来源）的函数。因此我们会发现，虽然实现数的一致性的条件得以满足，由于言语的环境不同，这条规则可能在一种场合实现了 75％，而在另

一种场合却只能实现 100%。这意味着实现这条规则的趋势应该体现在规则的形式标记上面，因为这也是说话人的语言能力的一部分，他知道“哪种频率在什么场合是合适的。”这样我们就可用 0 到 1 的数字把每一种环境范畴和语法的每一条有选择性的规则联系起来。数字表示的是规则应用的概率，这个概率是一个语言变量的语境和各种语言外部的参数（如说话人的年龄、社会地位，等等）的函数。

设 p 为一条规则在某一语言环境的应用概率，就有：

$$p = p_0 \times \alpha \times \beta \times, \cdots, + \omega \tag{8.1}$$

p_0代表了说话人的个别特征（如地区、社会、环境、语体，等等）所产生的概率值，而 α，β，…，ω 则是一些数值，表示变量的语言环境的那些特征的影响。如果这些特征表示为 A，B，… Z，那么 p 在下面的公式里就是：

$$p = p_0 \times \alpha(A) \times \beta(B) \times, \cdots, + \omega(Z) \tag{8.2}$$

即和特征 A，B，... Z，相关的参数 a，b，...，w 的数值所产生的应用规则的概率。而不应用规则的概率就是：

$$1 - p = (1 - p_0) \times (1 - \alpha(A)) \times (1 - \beta(B)) \times, \cdots, \times (1 - \omega(Z)) \tag{8.3}$$

在下面我们以语音为例，说明标记变量规则的结构的办法，这个办法也可用到句法分析：

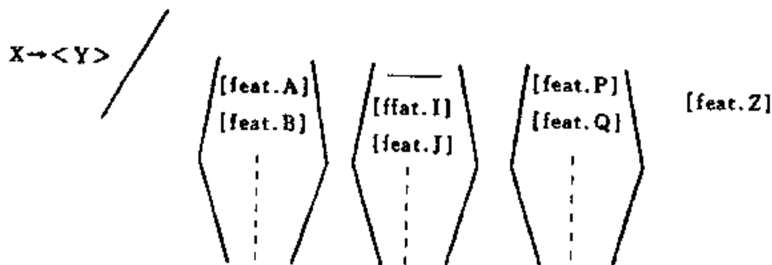


图 8.5 语音标记变量规则

根据变量约束（由箭咀括号表示）的数量权重的情况，X 可为 Y 所代替。每一对箭咀都包括有一些特征，在结构的描写中占有相当的位置。箭咀 > 和 < 相当与“多”与“少”，特征 Z 在方括号内，表示强制的特征，即规则的使用是强制性的。对由箭咀括号所表示的位置（前或后）来说，箭咀内清单所列的特征是互相排斥的，即在输入的字符串里，只能有一个特征。每一个特征只表示位置的一种可能性。

我们可以从语音、句法和语篇的不同的角度来看数据的形式化处理：

8.4.4.2. 语音

根据 Wolfram (1969) 在底特律对黑人英语的调查，美国非标准黑人英语在过去时态后

缀和辅音丛的发音中常有简化现象，在英语中以-t 和-d 结尾的辅音丛如 bold，find，fist，在非标准黑人英语中常念成 bol'，fin' 和 fis'。问题在于，这是辅音丛的简化？还是黑人英语中根本就少了最后的辅音？Labov 等人是这样考虑问题的：

(a) 没有哪个人从不使用这些辅音丛；也没有哪个人总是保留这些辅音丛；所以这是内在的差异。

(b) 对每一个人和每一个群体来说，第二个辅音的省略通常发生在接着的词由辅音开始的情况。

(c) 没有或很少有过度正确的情况。也就是说，错误的词类不会出现后面的-t 或-d，例如 mole 不会念成 mold，lip 不会念成 lipt。

这几点说明在 act，bold，find 这些底层结构里存在着完整的辅音丛，只不过有一条变量规则把第二个辅音省掉。下面一条规则覆盖了（a）和（c）三种情况：

$$t, d \rightarrow (\varnothing) \ /C —— \#\# \ (\sim V) \qquad (8.4)$$

这条规则的意思是：如果下一个词并非由元音开始，t 和 d 在一个辅音之后和一个词的界限之前可以有选择地略去。实际上很多中产阶级都说 firs' thing 和 las' month，而不说 firs' of all 和 las' October。但是这条规则没有概括（b），而在非标准黑人英语中，甚至在下一个词是由元音开始时，仍有 30 到 50%的时候是略去第二个辅音的。因此规则里应该反应这种情况，故规则可补充为：

$$t, d \rightarrow (\varnothing) \ —— \#\#\alpha \ (\sim V) \qquad (8.5)$$

α 视后面的环境是有元音还是没有元音而取正值或负值。但是非标准黑人英语中过去时态的后缀，也有同样简化辅音丛的情况，所以这条规则还要进一步补充。这里说明形成形式化的规则的方法，其余的就不再展开了。

8.4.4.3. 句法

在美国黑人本地语里，系动词“be”也往往是缩略或省略掉的，如：

(a) He a eat-and-runner.（名词前的 is 被省掉，表示为 [——NP]）
(b) He fast in everything he do.（形容词性谓语前的 is 被省掉，表示为 [——PA]）
(c) We are on tape.（在地点表达式前的 are 被省掉，表示为 [——Loc]）
(d) They not caught.（在否定式前的 are 被省掉，表示为 [——Neg]

如果单看这几个例子，我们可以下结论说，黑人英语本地语里是没有系动词的，但是情况没有那样简单，类似下面的用法也是存在的：

(e) She was likin' me.
(f) It ain't no can can't get in no coop.
(g) I'm not no strong drinker.

(h) Each year he will be gettin' worse all the time.

(i) Be cool, brothers!

Labov 对这种用法做了细致的分析后，认为首先应把缩略（contraction）和省略（deletion）区分开来：缩略是表达式的压缩，原来的句型还保留，如 He's here 是 He is here 的压缩；省略则略去原来句型中的一个部分，如 He is fast 被省略为 He fast。然后他提出下面几点看法：

1. 在黑人本地语中没有人完全省略系动词，也没有人完全不省略。每个人都有使用完整的形式、缩略的形式和不使用系动词的情况。这说明系动词用法可归纳为变量规则。

2. 在有些句法位置上不会出现省略：如省略句（He is too）和 wh-句后（如 That's what he is）。一般的情况是，标准英语可以缩略的地方，黑人英语可以省略；标准英语不能缩略的地方，黑人英语也不能省略。

3. 缩略和省略的这些关系使我们不得不去分析英语的缩略形式。我们发现 am，is，are，will，has，have，had 的缩略方式是移去包括有时态标记的词中的辅音前的单独的非重读的中元音（/ə/）。这个产生 He's here，I'm coming，You're here，I'll go，He's got it 的过程取决于一些省略起始的流音的规则和把元音弱化为非重读的中元音的省略元音的规则，而后者又取决于由句子表层结构的句法所决定的重音规则。

4. 系动词的省略和缩略有关还可以表示为另一种情况：黑人英语并不省略那些不能弱化元音的表示时态的形式，所以 be、ain't、can't 并不能缩略。省略是一个语音过程，例如在 I'm 中的 m 并不能省略：一般来说，后面的鼻音在黑人英语中不能省去。

5. 控制缩略和省略的变量规则按照语法环境的不同而显示出一系列变量约束。如果在前面的名词短语为代词，这个规则就会执行。而下面的语法环境则从最无力到最有力按次序对规则进行约束：名词短语作谓语、形容词和方位词、动词、动词前的助动词 gonna。如果我们认为先有缩略，然后才有把缩略后留下的单独的辅音省略的过程，那么这些约束对缩略和省略的两种规则所起的作用是相同的，而且黑人英语的缩略和其他方言的缩略型式是一致的。所以实际上是有两条规则。

6. 虽然同样的语法约束作用于黑人英语的缩略和省略，前置的元音或辅音的语音是刚好相反的。对缩略而言，如果主语以元音结尾，规则就会执行：对省略而言，如果主语以辅音结尾，规则就会执行。这种相反的情况和缩略与省略在语音上的差别是一致的，缩略移去元音，而省略移去辅音。两种情况都导致 CVC（辅音＋元音＋辅音）的结构的出现。

根据上面的观察，就可有两条缩略和省略的规则：

缩略规则

$$ə \rightarrow (\varnothing)' \begin{bmatrix} \alpha \text{pro} \\ \gamma \text{V} \end{bmatrix} \# \# \begin{matrix} ___ \\ +\text{T} \end{matrix} Z \# \# \begin{bmatrix} \beta \text{Verb} \\ -\gamma \text{Noum} \end{bmatrix} \quad (8.6)$$

省略规则

$$Z \rightarrow (\varnothing)' \begin{bmatrix} \alpha\text{pro} \\ \gamma\text{V} \end{bmatrix} \#\#[——]\#\# \begin{bmatrix} \beta\text{Verb} \\ -\gamma\text{Noun} \end{bmatrix} \quad (8.7)$$

Labov 的结论是在黑人英语里系动词的缩略或省略是在表层结构上出现的，其深层结构和标准英语是一样的。

8.4.4.4. 语篇

8.4.3.2 中第二段关于象牙海岸操 Abbey 语者进行问候的交际事件，也可用形式化的规则来概括：

问候次序→问候与反应＋请坐＋打听消息

(1) 问候与反应→

P2 问候 P1，如果　　P2 ＝
男性
＋成人
友善
a［女性］
儿童
－不友善

P1 反应，如果　　P1 ＝
成人
＋友善
a［较大的儿童］
－［不友善］

(2) 请坐　p1 请 p2 坐
P1 问候

(3) 打听消息

P1 向 P2 打听消息，　如果　P1＝
＋成人
男性
a［女性］
－［儿童］

P2 作出公式性的反应

这几个规则可以读成：

(1) 如果 P2 是友善的、男性成年人，他就会作出问候；如果 P2 是一个女性，她也会问候，但是不大可能；如果 P2 是不友善或是一个儿童，他就不会问候。如果 P1 是一个友善的成年人（男性或女性），他就会接受问候；如果 P1 是一个年纪大的儿童，他也可能接受；如果 P1 不友善，他就不会接受。

(2) 如果问候已经作出并被接受，P1（任何年龄的男性或女性）就会请 P2 就坐，并回报以问候。

(3) 如果 P1 是一个男性成年人，他就会打听消息；女性也会打听，但是不大可能；年纪

大的儿童虽然接受了问候，也不能打听消息。

这些规则是有序的：如果一条规则没有实行，其余的规则就不能执行。假定一个女性的P2并不向P1问候，P1就不能回应，不能请P2坐，也不能向P2打听消息。规则前面标有（+）的是总是执行的，（α）是可执行或不执行的，（－）是毫不执行的。

8.4.5. 语言态度调查

8.4.5.1. 什么是语言态度调查？

社会是由个人组成的，当一个人和另一个人交谈时，就会出现人们怎样对待语言，即语言态度的问题。研究语言态度牵涉到社会心理学。

关于语言态度有两种对立的观点：一种是心智主义（mentalist）的观点，它把语言态度看成是一种准备状态（a state of readiness），它是对人发生影响的刺激和该人作出的反应之间的一个中介变量。一个人的语言态度使他准备对某一刺激，而不对别的刺激作出反应。Williams（1974）对心智主义的语言态度观的定义是：态度是某一种刺激所引起的内部状态，这种内部状态对机体随后而产生的反应起中介作用。这种观点给实验方法带来问题，因为如果态度是一种内部的准备状态，而不是一种可观察的反应，我们就必须依赖一个人对他们态度的报告，或根据行为型式来推断语言态度。我们知道自我报告的数据往往并不很有效，根据行为型式而作的推断使我们更远离事实。所以很多语言态度的研究都要花很多精力来设计巧妙的实验，使被试能够在不意识到调查过程的情况下表示其态度。

另外一种是行为主义的观点：根据这种观点，语言态度只能来自人们对社会环境所作出的反应。这种观点不要求自我报告和间接推断，因而调查较容易进行；它只要对可观察到的行为进行观察、归类和分析。但是这种语言态度的调查不如前一种那样有意思，因为其结果不能用来预测其他行为。不过这种直接的行为主义的方法仍有它的用途。

和语言态度调查有关的另一个问题是语言态度下能否区分出一些子成分。一般来说，接受行为主义观点的人认为语言态度是不可分割的单位，而心智主义则认为语言态度还可分为认知的、感情的、意愿的子成分。

语言态度的调查和语言有关，例如向被试调查他们认为某一种语言变体是否“丰富”、“贫乏”、“美丽”、“丑恶”、“悦耳”、“难听”。这样一来，语言态度的定义往往被扩展到说某一语言或方言的人的态度，甚至其他的态度，如对语言维持和语言计划的态度。

从心智主义的观点来看，只要知道一个人的态度，就能准确地预测他的有关行为。语音变化的走向和言语社区是否赞成该变化有关系，甚至言语社区的定义都和大家的语言态度分不开。语言态度往往还反映了对某些民族成员的态度，或影响了教师对学生的态度，影响了第二语言的学习。有人还发现语言态度对一个人判断他是否听懂一种语言变体也有影响。

8.4.5.2. 语言态度调查方法

1. 直接法和间接法

完全直接的方法要求被试对调查其语言态度的问卷作出直接回应。完全间接的方法则不让被试知道别人在调查他的语言态度。Cooper & Fishman（1974）假设在以色列的人认为希伯来语是科学用语，而阿拉伯语在传递传统的伊斯兰教义方面更为有效。为了验证他们的假设，

他们找了一些懂得这两种语言的穆斯林成人，让他们听一些这两种语言的录音，一段是关于烟草害处，并提供科学证据的录音，用希伯莱语和阿拉伯语各录一次。另一段则是关于喝酒的害处，从传统的伊斯兰教义方面来支持，也用两种语言各录一次。听众分为两组，一组听用希伯莱语讲烟草的害处，用阿拉伯语讲喝酒的害处的两段话，而另外一组则听相反的搭配。结果出现戏剧性的差异：听用希伯莱语讲烟草害处的那一组比听用阿拉伯语讲烟草害处的那一组有更多的人说他们支持对烟草课以重税，其比例为2比1。而听用阿拉伯语讲喝酒的害处的一组比听用希伯莱语来说喝酒害处的一组刚好按相同的比例说他们支持对酒课以重税。假设得到验证，但是被试并不知道他们参加了一次语言态度的调查，他们的注意力集中在讨论烟酒的祸害。

2. 改变装束测试法（The match-guise technique）

Lambert 等人（1960）所建立的这种方法已成为语言态度调查的一种基本的方法。最单纯的改变装束方法的目的是控制语言以外的其他变量，首先是找一些能流利地说两种要观察的语言的人，让他们用两种语言读同一段话，并录下音来。在录音里故意安排成好像是两个人的录音。例如几个说话人用法语和英语各录一次音，录音带开始是一个人的一段法语，然后是另一个人的一段英语的装束，第三人的英语的装束，最后第四段的英语的装束是第一个人的。到这个时候，被试已经忘了第一个人的声音，就会以为这是另一个人的声音。其他人的录音也是用这种穿插安排的办法加以打乱，所以受试以为他听到的每一段话都是另一个人说的，他们以为自己所听到的话是实际上的话的两倍。

这些录音再拿来放给来自同一语言社区操双语的听众听，要他们对说话人的某些特征，如文化水平、社会阶层、相似程度进行评估，如果对同一个人的不同装束有不同的评估，其差异就是语言所做成的。因为同一个人提供了两种样本，对听众所产生的差异，不可能是音质差异所做成的。作为一个变量，内容也可以通过把同一段话翻译成另一文本而加以排除。改变装束测试法也算是直接法，因为它要求被试直接说出他们对某些特征的意见，但是它也有间接的成分，因为被试是在对说话人，而不是对语言作出反应，而且他们并不意识到他们听到的是同一个人的不同装束。

3. 语义微分量表（Semantic differential scales）

改变装束测试法常使用到语义微分量表（Osgood et al 1957）[①]，这些量表表示一个倾向的两端，中间有几个档次的空间，如：

友好 —— —— —— —— —— —— ——不友好
　　　1　　2　　3　　4　　5　　6　　7

让被试在量表上作出选择，然后加以汇总，再取得其平均值。这个平均值可用来作各种统计分析。

① 请参看10.7.3.3。

4. 问卷、访问、观察

(1) 问卷。问卷有开放式和封闭式两种，开放式使被试能够自由表示他们的意见，但他们答案容易偏离主题，难以打分。封闭式问卷有固定的格式，让被试填写，除了语义微分外，还有正误题、多项选择、排列等等。封闭式容易填写，也容易打分，但是它迫使被试按照调查者的意图去回答问题。比较好的妥协方法是先用开放式问卷了解反应，然后再根据反应来设计封闭式问卷。

(2) 访问。访问是没有问卷的开放式问卷。一个现场调查工作者提出有关语言态度的问题，让被试作出口头回应，然后再笔录或录音。被试无需在开放式问卷上作答，因此访问者可以较容易地提取问题，而且在被试离题时，能够及时防止。访问的主要问题是费时和花钱较多。作一次访问的时间比作 50 到 100 题问卷的时间还要多。

(3) 观察。这是收集数据的最自然的方法，宜于作人类学和民族学研究。从行为主义的观点来看，这也是最合适的方法，因为人们很少直接谈到自己的心理过程。从心智主义的观点来看，使用观察法必须进行推断，故难免犯主观主义的错误。

5. 改变装束测试法的问题和修正。这种方法有其内在的问题：

(1) 这种方法要求说话人用不同语言读同一篇材料，这就多产生一个变量，有可能从他读材料的角度，而不是从他使用语言的角度来对说话人进行判断，所以有人把内容改为让说话人讨论同一个题目，而不是读同一篇材料。但控制谈话内容又会产生另一问题，可能使语言变体和谈话内容不统一。如果在一个双语制的社会里，两个"装束"刚好是一个高变体，一个是低变体，对一种变体合适的题目不一定适合于另一变体。被试对一种装束评分较低，并非因为他们认为语言形式不合适，而是因为他们认为这种语言形式并不合适于讨论这个题目。

(2) 问卷的格式(包括改变装束测试和语义微分法)的效度问题。要验证认知和感情的态度几乎是不可能的，因为我们在比较的是实验结果和人们实际上所想和所感觉的东西。意愿的态度则较容易解决，因为它和行为有联系。Fishman (1968) 向纽约地区的波多利哥人调查他们对自己的民族性的态度，然后还举办一个波多利哥的跳舞晚会，请被调查的人参加。如果被调查人在问卷上回答他以作为波多利哥人为荣，而又参加晚会活动，他的回答就是有效的。

(3) 改变装束测试法的最后一个难题是和它的人工性有关。让人按照声音来判断人和实际生活很不相同。让被试去听同一内容的录音往往会使被试感到厌倦，所以才去更多地注意声音的变化。根据声音来在评分表上打分也不是日常生活常见的，Bourhis & Giles (1976) 设计了一个更为自然的改变装束实验，使被试不知道他们是在参加一个语言态度的调查。实验牵涉到威尔士地区的社会语言状况，当地有 4 种语言变体：英国英语中地位最高的 RP (Received Pronunciation)、略带南威尔士口音的英语、带较重南威尔士口音的英语、标准威尔士语。"被试"为戏院观众，而实验的刺激是中场休息时所作的通知，内容是要求观众填写一份问卷以帮助戏院安排以后的节目。"参加"实验的观众有两类：一类是居住在威尔士的英国人，只会说英语。他们有 5 个晚上每晚看两部英语电影。另一类为说双语的威尔士人，他们有 4 晚每晚看一个用威尔士语演出的戏剧。对英国人的观众所作的通知使用了 3 种英语变体，有两个晚上分别用 RP 和重威尔士口音英语，一个晚上用轻威尔士口音英语。说双语的威尔士人在

4 个晚上各听一种变体。所测量的行为两类听众听了不同变体的通知后的问卷回收率。数据是有还是没有回收，因此可用卡方检验[①]。结果表明：威尔士的英国人听了 RP 和轻威尔士口音英语后的回收率大致相等，一为 22.5%，一为 25%；而重威尔士口音的英语的回收率只有 8.25%。说双语的威尔士人的情况则刚刚相反：听了威尔士语后的回收率为 26%，而听了轻威尔士口音和重威尔士口音的英语的回收率各为 9.2%和 8.1%，RP 的回收率最低，只有 2.5%。

8.4.5.3. 语言态度调查的用途。

作为一种工具，语言态度调查有助于认识语言的社会作用。

1. 群体的认同。

语言对一个社会文化群体可以起到统一或分裂的作用。语言态度调查可以帮助我们了解群体的成员对某种语言的维系作用的认识，例如在希腊有一个叫做 Arvanites 的民族，而阿尔巴利亚语是这个群体所认同的语言。Trudgill & Tzavaras (1977) 对不同年龄的 Arvanitika 人进行了封闭式的问卷调查，了解他们对 Arvanitika 方言的态度。他们提出 3 个问题：

(a) 您喜欢说 Arvanitika 语吗？
(b) 您认为说 Arvanitika 是一件好事吗？
(c) 说 Arvanitika 是否有利？

调查结果如下：

表 8.11　　对 Arvanitika 方言的态度调查

	是			不表态			否		
年龄	喜欢	好事	有利	喜欢	好事	有利	喜欢	好事	有利
5—9	1	1	1	10	10	17	89	89	82
10—14	16	17	12	17	73	38	67	10	50
15—24	17	18	34	41	48	50	42	35	16
25—34	30	32	40	67	66	49	3	3	11
35—49	46	64	56	53	35	36	2	1	8
50—59	67	95	67	31	4	33	1	1	0
60+	79	86	97	21	14	3	0	0	0

结果显示不同年龄的人对 Arvanitika 方言的态度有很大的差异：年纪越大的人越喜欢 Arvanitika 方言。另外一个问题是“Arvanitis 人是否必须说 Arvanitika 语？”不同年龄组的人的回答是：

① 请参看 11.4.4.4. 的 χ^2。

表 8.12　不同年龄组对 Arvanitika 语的态度

年龄	是	否
10－14	67	33
15－24	42	58
25－34	24	76
35－49	28	72
50－59	33	67
60＋	17	83

这个数据好像和上表的数据所显示的型式不很一致：除了年轻人外，多数年龄组的人都认为 Arvanitis 人无需说 Arvanitika 语。Trudgill & Tzavanika 的解释是年龄较大的人意识到 Arvanitika 语正在消亡，但却希望保持他们的民族认同，因此也必须接纳那些不说 Arvanitika 语的 Arvanitis 人。而年轻的人对 Arvanitika 的前途没有希望，好像预见到语言和民族的消亡，他们的观点是只有说 Arvanitika 的才算是 Arvanitis 人，既然说 Arvanitika 的人越来越少，Arvanitis 人也就越来越少，但是这没有什么值得大惊小怪的。

2. 双语制

语言态度调查对双语制社区的语言型式可起到预测的作用。

El-Dash & Tucker (1975) 使用了改变装束的模型，他们找了两个既会说阿拉伯古典语和阿拉伯口语，又会说英语的埃及人来讨论敏感的 Giza 金字塔问题，而不是读一篇材料。他们说英语时自然会带有阿拉伯口音。这种带阿拉伯口音的英语处于阿拉伯古典语和阿拉伯口语之间的地位。然后他们使用了 4 个语义微分的量表：智力、领导能力、宗教性和喜爱程度，并把结果用来做方差分析①，发现语言变体在 4 个特征方面均有主要效应。但是光看主要效应并不能说明很多问题，例如不同装束在智力方面的平均分为：

表 8.13　不同装束被试的智力平均分

特征	阿拉伯古典语	阿拉伯口语	埃及式英语	美国英语	英国英语	F-比率	显著性意义
智力	10.25	8.49	9.59	9.13	8.45	20.33	p<0.01

表 8.14　双语制和其他特征的关系

特征	阿拉伯古典语		埃及式英语		阿拉伯口语
智力	10.25		*		8.49
	10.25		9.59	§	8.49
领导能力	8.74		*		7.20
	8.74		8.56	*	7.20
宗教性	9.38		*		7.75
	9.38	*	7.11		7.75
喜爱程度	9.48		*		8.51
	9.48	*	8.71		8.51

F-比率具有显著意义，这说明在 6 种装束里，某种装束的人会被判断为比其他装束的人具有

① 方差分析（Analysis of Varince，ANOVA），可参看 11.6。

更高的智力，但它并不能告诉我们哪一种装束的差别具有显著性意义。但是有一种 Newman-Keuls 复式比较的统计程序可以用来检查它们之间的差别是否有显著性意义。

* 表示在 $p<0.01$ 水平上两个分数有显著意义的差别

§ 表示在 $p<0.05$ 水平上两个分数有显著意义的差别

两个分数之间（如在智力方面，阿拉伯古典语和埃及式英语）没有符号表示它们没有显著意义的差别。

从表中可见，阿拉伯古典语和埃及式英语在智力和领导能力方面均高于阿拉伯口语，这和双语制的型式是一致的。虽然阿拉伯古典语的评分比埃及式英语略高，但其差别没有显著性意义。

3. 教育

除了了解社会结构以外，语言态度研究还可为教育提供信息，一般有两种类型：一是教师的语言态度，二是第二语言学习者的语言态度。第二种类型的研究是要了解学习者的态度是否影响他们的学习，这里不准备讨论。下面要讨论的是教师态度的调查和测量，以 Williams 在 70 年代中叶的研究最为突出。他的基本方法是让被试按照语义微分的量表来评估录下来的言语样本（录音或录像），其录音并没有使用改变装束的方法。但是所取的样本来自同一年龄的不同民族和社会阶层的儿童，录音为未加控制的自由交谈。在他的研究里，Williams (1974) 的主要精力放在设计语义微分量表，他采用了 4 个步骤：首先是试点研究，让少数教师听一些录音样本，然后用开放式问卷方法，让他们用自己的话来描写录音的说话人。然后他再从开放式问卷的反应中选择一些形容词，放到原型的量表里，目的是保证量表中所采用的形容词合适于教师使用。第三步是让另外一组教师使用量表来评估语言样本。最后是使用因子分析的统计手段来显示反应的基本的维度。

Williams 研究的结果发现了能够有效而又可靠地表示两维以上的教师态度的一个二因素模型。一维是信心和热切程度，它测量教师根据儿童的话语流利程度和积极性而作出的整体反应的态度；另一维是民族性和非标准性。它反应教师对不同社会阶层（如高阶层和低阶层，白人和非白人）的语言特征的态度。他使用回归[①] 的统计手段说明实际言语样本中的一些特征可以用来“预测”这两个维度的结果。例如一些非标准的语法或语音特征（把 d 发成 th）的频率可以预测民族性和非标准性量表上的得分。换句话说，如果一个儿童经常说 dem 和 dose，在民族性（在 Williams 的研究里，指美国黑人或美国墨西哥人）和非标准性量表里，得分就会高。同样，uh，uhm 这种“口吃现象”也能成功地预测在信心和热切量表上的得分。这个研究的含义是教师的反应根据两个因素：儿童的表达的形式（民族性和非标准性）和他们说话的方式（信心和热切程度）。

8.4.6. 语言用法调查

上一节所谈的是从宏观社会语言学的角度看社会对语言的态度，下面要谈的是社会对一些语言用法的态度。这里牵涉到用法和态度两个方面：一是被调查人自己是否这样或那样用，

① 指在相关系数基础上所作的线性回归分析，请参看 11.5。

二是他对这种或那种用法的态度如何，是赞成还是反对？还是虽然不赞成，但能容忍别人这样用？一般来说，个人的用法和他对这种用法的态度是一致的，但是也会出现不一致或不完全一致的地方。

8.4.6.1. Fries 的《美国英语语法》

描写语言学历来对用法的调查都很感兴趣：Jespersen 在主编他的《现代英语语法》时，就采取了历史主义的原则，对一些有争议的用法都从正反的角度加以客观的描写：什么时候哪些作家是这样用的，哪些作家不是这样用的。Fries（1940）在主编《美国英语语法》时以美国政府提供的 3000 封信件（其中完整的 2000 份，摘录的 1000 份）为依据，描写美国英语的用法。这些信件都出自已居住在美国三代以上的美国人。Fries 按照写信人的社会阶层，把资料分为三类：（Ⅰ）标准（standard）英语，写信人为起码在大学念了三年的大学毕业生，从事各种专业性工作；（Ⅱ）普通（common）英语，写信人受过从高中一年级到大学一年级的教育，从事间于专业性工作和体力劳动之间的工作；（Ⅲ）通俗（vulgar）英语，写信人所受的教育不超过中学八年级。有的接近文盲，从事体力劳动工作。Fries 主要比较（I）和（III）在用法上的差异。他认为比较不能凭印象，那样容易导致只注意差异，而忽略那些一致的地方。所以他把观察范畴内所有的语言事实都记录下来，然后统计出差异的相对频率。他的比较主要从三个方面进行：

(a) 词的形式。

(b) 功能词

(c) 词序

限于篇幅，我们在这里只举一个例子，以兹说明：

表 8.15　　英语中的 have＋过去分词用法

	标准英语		通俗英语	
have＋及物动词的过去分词：				
have＋had	39	8.8%	16	7.8%
have＋been＋过去分词	88	20.0%	16	7.8%
have＋其他及物动词的过去分词	232	52.5%	104	50.3%
小计	359	81.2%	136	66.3%
have＋不及物动词的过去分词：				
have＋been	49	11.0%	54	26.4%
have＋其他不及物动词的过去分词	34	7.8%	15	7.3%
小计	83	18.8%	69	33.7%
总计	442		205	

Fries 发现下列有趣的事实：

1. Have＋过去分词在标准英语的使用频率远远高于它在通俗英语的频率。过去分词在标

准英语的出现率（1157）为通俗英语的出现率（311）的 4 倍，标准英语的 Be＋过去分词为通俗英语的 6 倍（712：117）。但是标准英语的 have＋过去分词的出现率仅为通俗英语的两倍（442：205），这说明通俗英语比标准英语更为保守。因为在古英语里，have 只与及物动词的过去分词连用，至于不及物动词，be 特别与表示动作的动词连用。但到了早期中古英语，have 首先是和不带宾语的表示动作的动词连用，然后和不及物动词，特别是 be 连用。

2. 两类英语中的一些例证的分布也特别有趣。有 3 种情况是相同的：(a) Have＋had 在标准英语中为 8.8%，在通俗英语中为 7.8%；(b) Have ＋其他及物动词的过去分词在标准英语占 53.4%，在通俗英语为 50.7%；(c) Have＋其他不及物动词的过去分词在标准英语为 7.8%，在通俗英语为 7.3%。

3. 这两种英语在两个方面是有显著差别的：(a) Have＋been＋过去分词在标准英语为 20%，在通俗英语为 7.8%，(b) Have＋过去分词 been 在标准英语为 11%，在通俗英语为 26.4%。

8.4.6.2. Mittins 的英语用法调查

Mittins 等人（1970）在英国所作的用法调查采取了不同的方法，也值得介绍。他们把历来语法学家有争议的一些用法集中起来，共 50 个，请了 457 人（教师为主）从非正式口语，非正式笔语，正式口语，正式笔语 4 个方面来判断这些用法在哪些方面是可以接受的。这些有争议的用法有下面的 5 种类型，其接受的百分比分别为：

1. 口语项目，有 8 个，如：under these circumstances；they will send... providing the tax is low；should try and arrive 等，接受率 40%。

2. 词源项目，有 5 个，如：very amused；data is；these sort of plays need 等，接受率 55%。

3. 语法项目，有 18 个，如：at university；we have got to finish，he is older than me；is very different to 等，接受率 38%。

4. 词汇/语义项目，有 12 个，如：ended early，due to illness；everyone has their off-days；neither author nor publisher are 等，接受率 38%。

5. 语言神话，有 7 个，如：did not do as well as；his family are；told Charles and I 等，接受率 45%。

8.4.6.3. Quirk 等人的现代英语用法调查

规模巨大而有影响深远的是 Quirk 等人所主持的英语用法调查，其结果是不但建成了上千万词的语料库，而且资料还被利用来编写了几本重要的语法，如：《当代英语语法》(A Grammar of Centemporary English)，《英语语法大全》(A Comprehensive Grammar of the English Grammar)，《交际英语语法》(A Communicative Grammar of English)，《大学英语语法》(A University Grammar of English)。Greenbaum & Quirk (1970) 还专门编写了一本介绍他们所使用的调查方法的专著：《英语使用和态度研究中的提取实验》。书中首先用一个图来表示用法和态度的关系：

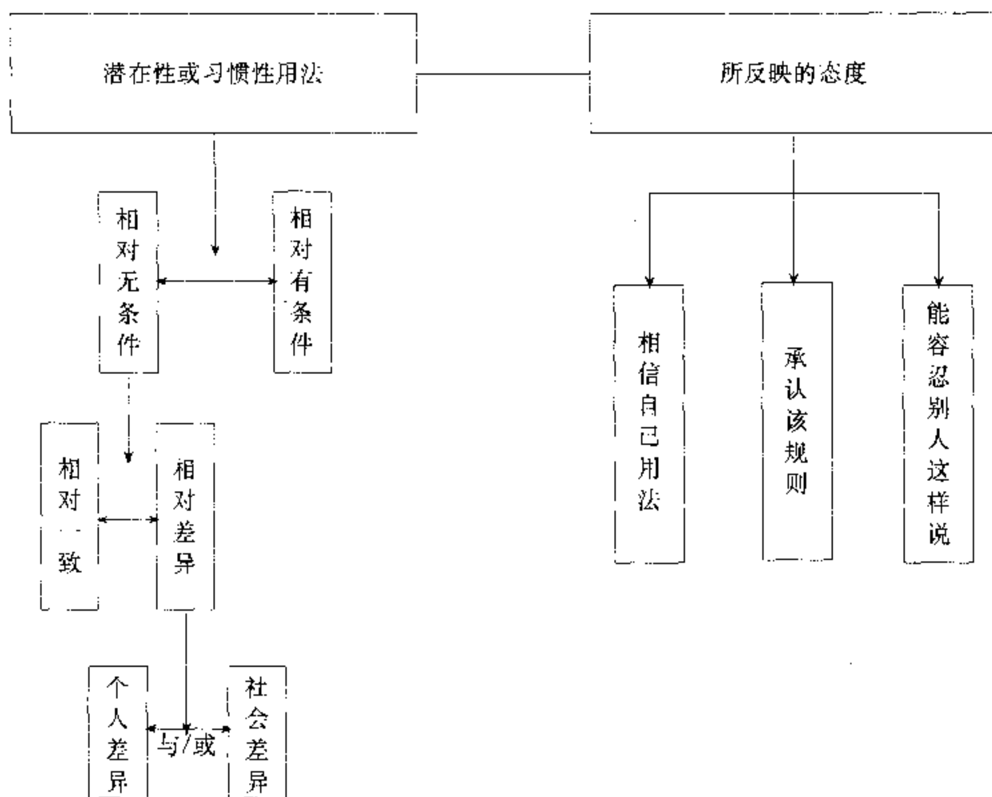


图 8.6 用法和态度

图中并没有“实际的”使用，因为实验的目的是超出（像语料库所记录的）实际的用法，而集中在考察提取技巧所需要的材料。但是对所提取的“潜在的”和“习惯性的”材料必须作出区别：一种情况是所提取的句子所体现的主要特征是被试以前碰到过的（例如 learn 的过去式，或 hardly 出现在助动词与非限定动词之间的用法），这就是习惯性的；另一种情况是所提取的句子所体现的主要特征是被试从来没有使用过的，但它是他的语言能力的范围里面的，这就是潜在的，例如要求被试提供一个奇怪的动词/flaiv/的过去式，或是看被试把 introductorily 或 blondely 放在什么位置。图中还区分“有条件”和“无条件”两种用法，前者意味着“受特定的语言或环境因素所限制”，后者指不受限制。这两者体现了一个量表的两端。同样的，“相对一致”和“相对差异”也是一个量表的两端，后者可理解为“自由变异”。差异是否能够完全无条件的，这是值得怀疑的。这种差异可能是个人的特征，例如某人把 psychology 的 psy-有时念为/sai/，有时念为/psai/，这种摇摆并不反映整个社会的情况。又如 either 的 ei-在社会上有两种读法，有时念为/i：/，有时念为/ai/，这种情况并不反映那个把 ei-念成/ai/的人的也有两种发音。当然这两者并非互相排除的，社会上的差异往往和个人的差异相一致，因为社会是由个人组成的。

所提取的用法和对这些用法的态度也应有所区别。这些态度反映了 3 种不一样的，但又相互有联系的因素。我们可以对我们习惯使用的形式有强烈的信念，也可以对应该使用的形式有强烈的观点；这两点可以是一致的，但也可能是有冲突的。但我们对自己用法的信念不一定和我们实际的用法相一致，这也是无庸置疑的。而且我们还可容忍别人使用一些和我们自己的用法不一样的或和我们的信念不一样的用法。（图 8.7）进一步说明他们所使用的两种不同的测试形式：一种是用法测试，另一种是态度测试。

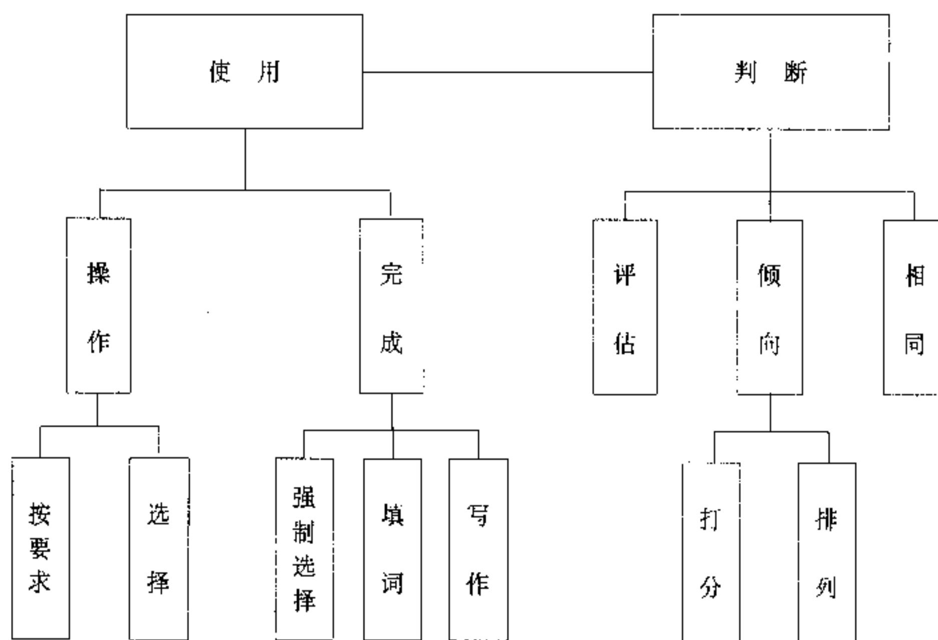


图 8.7　用法测试和态度测试

用法测试（Performance tests）包括操作（Operation）和完成（Completion）两类测试，对被试有不同的要求。

操作测试要求被试对所给的句子作一些改变，它又有两种：

(a) 按要求进行（Compliance）的测试，它让被试听一些违背规则的句子，要求他们改变其中的一些词，然后观察改变后的句子有没有把违背规则的地方改过来，例如让被试听 He hardly could sit still 后，要求他们把 he 改为 they。测试者预测被试不会接受 hardly 在句子中的位置，因此被试在改变 he 时会不自觉地把 hardly 放在他们认为合适的地方，在助动词和动词之间，说成 They could hardly sit still。

(b) 选择（Selection）测试，听的是正常的句子，但是当被试进行操作时，他们就面临选择（不管他们自己是否意识到），例如要求被试把 None of the children answered the question. 的句子里的时态改为现在时，被试就要决定用动词的单数还是复数。

完成测试要求被试对所给的句子增加一些东西，它有 3 种：

(a) 强制选择（Forced-choice selection）测试，用以了解不同的用法。实验向被试提供有限的选择项和有限的使用该项目的环境，例如要求被试在下面的环境里选择使用 learned 和 learnt：

I ____________ the poem.

I have ____________ the poem.

在这个测试里，所了解的不仅是被试倾向于使用哪一种形式，而且作为过去式或过去分词时使用哪一种形式。

(b) 填词（Word-placement）测试．它用来考察词的位置。它向被试提供一个句子和一个必须和句子一起使用的词，例如 My brother plays the guitar 和 usually，要求被试在句子中使用这个词，然后看被试把它放在什么位置。

(c) 写作（Composition）测试。它是开放性的，让被试听的是句子的一部分和这一部分

在句子中的位置，然后要求被试随意完成句子。例如听的部分放在句子开始的 I entirely，实验者感兴趣的被试会连用什么动词，以比较在别的情况下（如 I completely）被试会用什么动词。

判断（Judgment）测试有 3 种：

(a) 评估（Evaluative）。一般和按要求进行的测试一起使用，要求被试在一个三点量表上评估："完全自然和正常"、"完全不自然和不正常"，"在两者之间"。例如让被试评估以前做过的 He hardly could sit still 是否可以接受。

(b) 倾向（Preference）。一般和选择测试一起使用，包括两个部分：打分和排列。例如让被试听一个句子的两种变异的形式：None of the children answers the question 和 None of the children answer the question，然后让他们在上述三点量表上打分。由于向被试提供两种形式，他们的注意力就会被吸引到变异点上面。最后还要求被试按他们的意见来排列两者的先后。

(c) 相同（Similarity）。也牵涉到对句子关系的判断，但主要是语义关系。测试向被试提供的是两个句子，其中有最小的语义或句法差别，然后要求被试在一个三点量表上评估它们是否相同："意义很相同"，"意义很不一样"，"在两者之间"，例如：

/some lectures are actually given before ten#

/actually#/some lectures are given before ten#

Greenbaum & Quirk 认为他们实验的目的是了解接受性的各个方面的问题，但是关键的不是一种形式可接受的或不可接受，因为接受程度是分级的。他们关心的不能接受的程度有多大，更准确地说，在哪一点上不能接受。这可以从被试对违反规则的句子的"修正"方向看出来，例如被试听到的是 The council lowered his rent slightly，指令是把动词改为现在时，哪些移动 slightly 的位置（如念成 The council slightly lower his rent）的被试就能比哪些不改变位置的被试提供更明确的用法信息。而且在那些在评估测试中认为这个句子是可接受的被试中，有些人在按要求进行的测试中把副词的位置也移动了，这说明态度上的容忍和使用上的倾向性是不一致的。而且使用上的倾向和态度上的倾向也不一定一致的，例如 None of the children 后面跟着的是单数动词，这一点在评判测试中主张的比在选择测试中主张的要多。

整个实验包括 50 道使用测试题和 50 道判断测试题，刚好一堂课可以做完；被试全是大学生。实验者站在被试面前，用口头方式（后来改进为使用录音机）念出试题，指令都是事先写好的，以保持一致。指令并没有说明所听的句子中有些是不符合语法的，而只是提出任务，举出例子。操作测试要求被试完成任务有：把时态改为现在时或过去时；否定式和肯定式互换；把做主语的代词改为单数或复数的形式；把陈述句改为由 be 或 do 引导的问句。一直到判断测试题前，实验者都不让被试知道测试是和调查语言用法有关。到了判断测试题，他们才知道他们"将会听到同样的题目，时间较短"，但要在一个三点量表上判断其可接受程度：

完全自然和正常

在两者之间，值得怀疑

完全不自然和不正常

英语用法的范围很广泛，实验者决定先从调查副词的位置做起，大部分试题都与此有关，后来才决定增加一点句法和词汇的项目，但数量不多。

实验收集了各种各样的数据，怎样对答卷评分，也是一个值得考虑的问题。判断测试是直接了解态度的，这比较好办；但是用法测试则不同，出现了很多情况，以按要求进行的测试为例，可以有不同的回应。例如要求把 he will/probably stay late# 改为问句，期望的答案是 Will he probably stay late? 测试的目的是了解副词的位置：

（1）完全按要求做，产生实验者事先期望的句子（A）

（2）犹疑，有 3 种不同情况：（B）在所要了解的问题以外产生犹疑，例如回应为 Will he probably ‹stay› late? 三角括号表示有犹疑，后来才插进去的；（C）是所要了解的问题的中心，但与回避无关，例如对 probably 的拼音没有把握产生犹疑；（D）和回避有关，例如 Will he ((probably)) stay late? 双圆括号表示产生犹疑，把 probably 删去。

（3）没有按要求做，答案有 4 种不同情况：（E）所做的改变与所要了解的问题无关，例如答案为 Will you probably stay late?（F）是所要了解的问题的中心，但与回避无关，例如答案为 Will he stay late probably?（G）是所要了解的问题的中心，有回避，例如答案为 Will he stay late? 回避了所要了解的 probably 的位置的问题；（O）完全略去。

在这些不同的回应中，（G）和（O）是最足以测量没有按要求做的指标（RNC，the most relevant measure of non-compliance）。在统计中，RNC 的分数被用来说明被试没有按要求做的情况有多少。下面是两个例子：

例 1

		RNC 分	
		N＝85	%
E13 he /hardly could sit still#	he→they	54	63
E14 he could hardly sit still#	he→they	1	1

这个结果可以拿来和判断题的结果比较，看实际用法和态度是否一致：

	N＝85			%
	＋	?	－	'－'
E13 he /hardly could sit still#	16	23	46	54
E14 he could hardly sit still#	85	0	0	0

两者结果相当一致，认为 E13 '完全不正常' 的有 54%，而改动了 hardly 位置的有 63%，相差没有超过 10%。下面的一个例子也是想了解副词位置的，就没有那么一致：

		RNC 分	
		N＝85	%
E9 he /virtually is ruling the country#	he→they	65	76
E10 he is /virtually ruling the country#	he→they	5	6

但判断题的结果为

	N=85			%
	+	?	—	'—'
E9 he /virtually is ruling the country #	19	27	39	46
E10 he is /virtually ruling the country #	84	1	0	1

就态度而言，46%的人认为不能那样说，而在实际用法中有76%的人对virtually的位置作了移动。下面是Greenbaum & Quirk所给出的一幅态度和用法的比较图（不完全）：

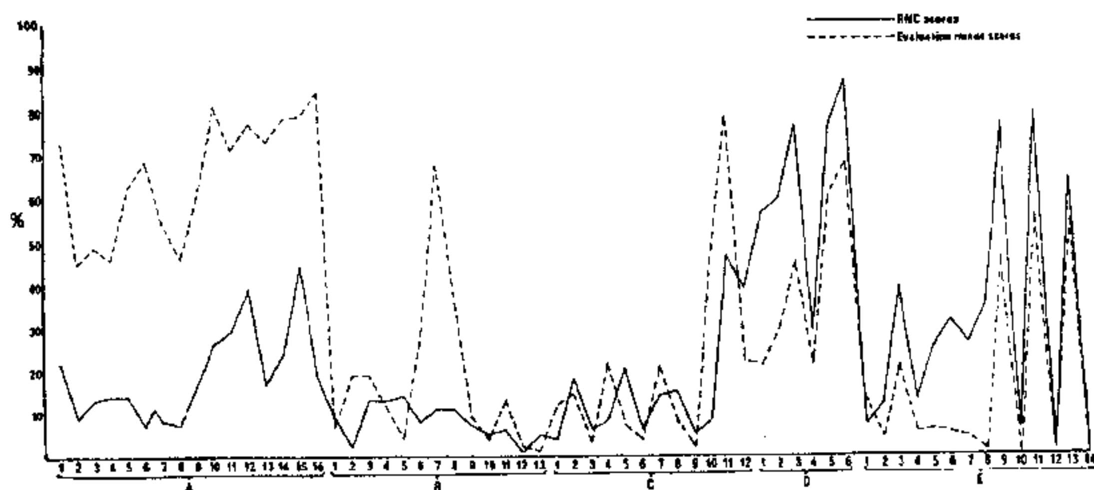


图8.8 用法和态度的比较图

8.4.7. 定量分析

上面所介绍的方法虽然也用到了一些定量的手段，但基本上是定性分析。使用这种方法的目的是发现语言在社会中的功能，决定一种语言必须具有什么特征才能完成一种功能。下面是根据Fasold（1984），介绍使用定量方法来研究社会的多语制。

8.4.7.1. 人口普查和调查

这是一种可靠的数据收集途径，但也应注意它的问题；

1. 问卷设计。与语言有关的问题有3类：(a) 母语的问题，企图了解被调查人的第一语言状况；(b) 使用的问题，企图了解经常使用的语言的状况，提问可以和语境有关或无关；(c)能力问题，企图了解被调查人对自己说话能力的评估。但是各国人口普查中所提的问题的内容和方法都不同，不易作比较。例如就母语所提的问题和就使用所提的问题就不好比较。甚至问的是同样性质的问题，用词不同也会引起问题：1940年在巴西普查中所提的问题是“你能流利地说民族语吗?”在墨西哥所提的是“你能说民族语吗?”回答后一个问题的人会比前一种的多得多。有的问题过于含混，也难获得可靠数据，如：“你懂哪一种语言?”有的普查的用词往往和上一次普查不一样，也难作纵向研究，如在巴拉圭，1950年问的是能力问题：你

能说哪些语言？在1962年问的是使用问题：你习惯使用哪些语言？如果一个人懂得西班牙语和Guarani，他在1950年就会把两者都列上，但这1962年，却只会回答一种。在印度1961年的普查中没有把懂得标准化的母语和相关的一种方言的人算成是双语使用者，但在1971年的普查里却没有这条规定。在美国1920年和1940年的普查中，对母语的定义都不一样：在1910和1920年间，第二代移民是按照其父母的母语来分类的；在1940年则是按在家里或从小说的语言来分类。政府所设计的问卷中，还往往按国家的民族语言政策来提问的，对本地语就会有歧视。例如在墨西哥，一个说西班牙语和另一种民族语言的Amerindian人说成是双语使用者，但是一个会说两种Amerindian语的人就不算。有时“语言”和“方言”的界限不易划分，也会引起混乱。在奥地利普查中，说Yiddish语的人被算成是说德语的人，以增加某些省份说德语的人的数目。在匈牙利却相反，南部斯拉夫少数民族是按其地方方言来分开统计的，以减少塞族——克罗地亚族人的数目。

2. 反应。“语言”和“方言”的界限难以区分，而且对第二语言的掌握程度上有很大的差别，所以当一个被调查的人对普查官员说他懂得一种语言时，很难断定他指的是什么。在印度北方，语言的流动性很大，所以当一个印度公民要回答他的母语和第二语言是什么时，任意性很大。另外对自己语言能力的评估，也有很大差异。在巴拉圭，掌握西班牙语是一种地位的象征，所以很多人都说他们懂得西班牙语。如果事先知道这种情况，就可以提出一些纠正手段。有时，语言问题会被人误解，例如在1961年的加拿大普查中，一些母语为英语的人把“你会说法语?”理解为“你懂法语?”

3. 地理。地理因素在两个方面影响普查：一是自然地理；二是地理边界。在有些国家里，由于山脉、森林妨碍交通，某些地区没能进行普查。更为复杂的问题是按照地理边界来对数据进行归类，在印度Trivandrum地区，语言差异指数为0.48，这意味着如果随机抽两个人，他们说的不是一种语言的可能性为50%。但实际的情况是该地区的东部的居民都是说Tamil语的，西部的居民则说Malayalam语。不过他们很少交往，不存在听不懂对方的问题。

4. 数据处理。普查的数据的收集、编辑和计算也有不少问题。主要问题出自现场记录者，数据放在哪一个框架里，全靠他们做主。在1961年印度的普查里，对他们的指示是列举被调查者除母语外能说的两种语言，但如果是母语的一种方言，就不算为第二语言。在指令里，并没有说明什么语言算是方言，这就引起归类的混乱。

尽管存在这些问题，普查的数据仍被广泛使用。因为它们是唯一能够大规模收集到的数据，而且差不多是在同一时间收集。普查数据还是唯一的可以进行跨国比较的数据。不过在使用这些数据时，必须注意：

1. 效度检验。一种是外部证实(external verification)，另一种是内部一致性(internal consistency)检验。外部证实较有说服力，我们必须找到其他数据来验证普查数据，例如在加拿大魁北克省的普查里，调查了母语为法语和母语为英语的数字，刚好在那里实行的是罗马天主教和基督教的学校制度，几乎所有的基督教学校都用英语作为授课语言，而天主教学校则使用法语。所以说法语的儿童上天主教学校，而说英语的儿童则上基督教学校。所以这两种学校的招生入学的数字应该接近于在学校读书的儿童的母语数字，Lieberson (1967) 举出下表：

表 8.16　学校招生和普查数据（10—14 岁儿童）的比较

学校招生	百分比	母语	百分比
罗马天主教	89.6	法语	85.9
基督教	10.4	英语	11.3
		其他	2.8

两组数据十分接近，说明普查反映了实际情况。在天主教学校的儿童的百分比高于以法语为母语的儿童的百分比，也是可以解释的：在天主教学校制度里的英语学校多于基督教学校制度里的说法语的学校，所以在天主教学校读书的说英语的儿童多于在基督教学校读书的说法语的儿童。

内部一致性检验可在两次以上的普查之间进行：如果在一次普查里报告某一年龄档中有多少人是使用某一语言为母语的，在 10 年后的普查里，比该年龄档大 10 岁的人中应该有相同比例的人使用该语言为母语。但是这必须有几个前提：(a) 所比较的年龄组的死亡率应该大致相同；(b) 在这 10 年间不会发生某一年龄组中有失去比例的人离开该国；(c) 在这 10 年间不会发生某一年龄组中有失去比例的人移民到该国。如果移民都有登记，第三项可以控制。Lieberson 对 1951 年加拿大普查中关于以英语或法语为母语的人的统计作过一次内部一致性的分析：他比较了 7 个年龄组，例如在 1951 年 10-14 岁年龄组中有 57.0%的人说英语，33.8%的人说法语。到了 1961 年，20-24 年龄组中有 56.8% 的人说英语，34.1%的人说法语。对 7 个年龄组的统计表明，说英语的人的差异为 0.8%，说法语的人的差异为 0.1%。这说明普查结果是很一致的。

2. 从普查中进一步提取数据。McConnell (1979) 先统计一个国家里以语言 X 为母语的人的总数，然后统计以其他语言为母语的人把语言 X 作为第二语言而学习过的人的总数，把两者加起来就是说语言 X 的人的总数。Lieberson 则从以英语为母语的人数中减去只会说英语，但不会说法语的人数，就可得到以英语为母语，但又会说法语的人数。

3. 推断。如果有足够的背景资料，我们也可从普查数据中进行推断。McConnell 根据土耳其 1965 年的普查作过一次推断：土耳其语是官方语言，但 Kurd 族在土耳其为数不少，说土耳其语的人在数量上和政治上均占优势，所以 40%的 Kurd 族人都把土耳其语作为第二语言而学习。有 40 万的说土耳其语的人在普查中登记他们的第二语言为 Kurdish，这些人不大可能是土耳其人，他们本身是 Kurd 族人，但在语言上效忠于土耳其语。这可用来说明 Kurd 族人的语言转移。

8.4.7.2. Greenberg-Lieberson 公式

社会语言学定量分析的核心是测量语言差异，以反映多语制的状况。最精细的方法是 Greenberg 所提出，后又为 Lieberson 所发展的公式。Greenberg (1956) 共提出了 8 条公式，第八条是“交际指数”(index of communication)。其他 7 条都是与语言差异有关，一条比一条精细。第七条最准确，也最复杂，而且所要求的数据往往不易得到。下面介绍的是第一条，也是最简单的一条。

Greenberg 假定我们把一个地理政治单位的所有的人放在一个口袋里，随机抽取两个人，而这两个人又都有交际意图，我们就必须了解他们是否共享一种语言。例如在一个 10 万人都说 X 语言、而有没有人说其他语言的“口袋”里，要算出抽到说 X 语言的人概率，就必须把说 X 语言的人数除以总人数，即 100 000/100 000=1。我们绝对有把握抽到一个说 X 语

言的人。"口袋"里现在还有 99 999 人，在 10 万人中少了一个人是微不足道的，所以概率仍为 100 000/100 000=1，要计算两个人都说 X 语言的概率，就必须把这两个人说 X 语言的概率相乘，即 1×1=1。这是一个没有语言差异的地区。

相反的是完全有差异的地区，在 10 万人中有 10 万种语言，即 10 万人中只有一个人说 X 语言，抽到说 X 语言的人的概率为 1/100 000，抽到另一个说 X 语言的人的概率为 0/100 000。1/100 000×0/100 000 =0。

Greenberg 所举的是一个假想的例子：在一个地区有 1/8 的人说 M 语，有 3/8 的人说 N 语，有 1/2 的人说 O 语。如果要抽两个说 M 语的人，其概率就是 1/8×1/8，即 1/64；要抽到两个说 N 语的人的概率就是 3/8×3/8=9/64。说 O 语的概率最高，1/2×1/2=1/4。如果我们要知道两个人说任何一种相同的语言的概率就必须把这 3 个概率加起来：1/64+9/64+1/4=26/64。

但是我们的最终目的是测量语言差异。语言差异的意思是随机抽取的两个人说不同语言的概率。这刚好和上述事例相反。如果原始数据方便，我们当然可以采用直接的方法来计算差异性，但是一个更为简单的方法是用 1 来减取两个人说相同语言的概率。在上面最后的一个例子里，说任何一种相同的语言的概率两个人为 26/64，因此剩下的就是不说同一语言的概率，即 1-26/64=38/64。转换为分数就是 0.594。

我们可用公式来表达上述的计算，用 i 来代表上例中的 3 种语言，1/8×1/8 可写为 $(1/8)^2$，便有说同一种语言的概率：

$$\sum i(i)^2 \tag{8.8}$$

用 1 来减去它，便有语言差异的公式，这就是公式 A：

$$A = 1 - \sum i(i^2) \tag{8.9}$$

公式 A 是从几个不现实的假定出发的：(a) 从一个政治地理单位里的任何地方抽取的两个人都同样有需要进行交际；(b) 说任何一种语言的人都完全不懂得另一种语言；(c) 不存在有任何说多种语言的人。

Greenberg 的第八条公式 H 是交际指数，测量的是交际，而不是语言差异。在公式 H 里，不需要用 1 来减去所求得的值。它和公式 A 的另一个不同的地方是，它需要考虑说多种语言的人的情况。它的基本的计算方法是：(a) 像公式 A 一样随机抽取一对说话人；(b) 注意那些能够进行交际的，即共享一种语言的说话人；(c) 计算抽取到每一对"好的"说话人的概率；(d) 把这些概率累加起来。例如，在某一个国家里有 3 种语言 M，N，O。在全体人民中说这几种语言的情况是：

说 M 语的	15%
说 N 语的	20%
说 O 语的	5%
说 MN 两种语的	25%
说 NO 两种语的	30%
说 MNO 三种语的	5%

第一步是列举 M 和 M，M 和 N，M 和 O，M 和 MN 等等说话人的对子如下：

表 8.17　　一个国家内说三种语言的概率及交际指数的计算

	M	N	O	MN	MO	NO	MNO
	(0.15)	(0.20)	(0.05)	(0.25)	(0.00)	(0.30)	(0.05)
M (0.15)	0.0225			0.0375	0.0000		0.0075
N (0.20)		0.0400		0.0500	0.0000	0.0600	0.0100
O (0.05)			0.0025		0.0000	0.0150	0.0025
MN (0.25)	0.0375	0.0500		0.0625	0.0000	0.0750	0.0125
MO (0.00)	0.0000		0.0000	0.0000	0.0000	0.0000	0.0000
NO (0.30)	0.0450	0.0600	0.0150	0.0750	0.0000	0.0900	0.0150
MNO (0.05)	0.0075	0.0100	0.0025	0.0125	0.0000	0.0150	0.0025

在表中有些是空格，如 M 和 N 交叉的空格，因为只会说 M 语的人不可能和只会说 N 语的人交谈。有可能的配对的交际的概率仍按上面提到的（8.8）公式计算，把两种概率相乘。第二步在把这些概率累加起来，得 0. 8350，就是这个国家的交际指数。

8.4.7.3. Lieberson 的延伸

Lieberson（1964）进一步延伸了 Greenberg 的公式。Greenberg 只是从地理的概念来考虑，但是交际指数可以延伸到任何得总体，如对罗马天主教徒的调查。更重要的是他的公式只是指一个总体内部的交际指数，并没有接触到两个母体之间的交际情况。例如在阿尔及利亚 1948 年的调查，其交际指数为 0.669，这意味着如果随机抽取两个阿尔及利亚人，他们共享一种语言的可能是 2/3。但是这个指数没有说明如果先抽取一个欧洲人，再抽取一个穆斯

表 8.18　　1948 年阿尔及利亚人口中跨民族交际指数表

欧洲	穆斯林					
	阿拉伯语	Berber 语	法语/阿拉伯语	法语/阿拉伯语/Berber 语	法语/Berber 语	法语/法语
	(0.780)	(0.112)	(0.053)	(0.047)	(0.007)	(0.001)
法语/阿拉伯语 (0.135)	0.105		0.007	0.005	0.001	0.000
法语/阿拉伯语/Berber 语 (0.003)	0.002	0.000	0.000	0.000	0.000	0.000
法语/Berber 语 (0.001)		0.000	0.000	0.000	0.000	0.000
法语 (0.860)			0.046	0.040	0.006	0.001

林，他们能否交际？Lieberson 还指出一种极端的情况：如果一个国家有两个群体：A 和 B。所有 A 的成员只会 X 语，所有 B 的成员只会 Y 语。整个总体的交际指数便不可能低于 0.500，如果其中一个群体的成员特别多，指数还会高。但是两个群体其实并不能交际。Lieberson 把 Greenberg 的公式略为改变，在阿尔及利亚人口的总体里先分出穆斯林和欧洲人，然后统计其交际指数。

按照这种算法，交际指数只有 0.214。这种延伸也可适用于 Greenberg 的公式 A，因此它可改写为：

$$A = 1 - \sum i(i1)(i2) \qquad (8.10)$$

i1 为一个群体中说某一种语言的人的比例，i2 为另一个群体中说同一语言的人的比例。例如挪威籍外国白种人在美国的母语分布（1940）情况是：

表 8.19　　挪威籍外国白种人在美国的母语分布

	母语		
	挪威语	英语	瑞典语
不在美国出生	0.924	0.051	0.012
第二代	0.502	0.477	0.008

其语言差异就是：

$$1-((0.924\times(0.502)+(0.051)\times(0.477)+(0.012)\times(0.008))=0.512$$

这个指数的含义是：如果从两代人中随机抽取两个人，他们不共享一种语言的机会为 50∶50。

8.4.7.4. 定量方法在广义社会语言学中的应用

定量方法在广义社会语言学里的典型的应用是研究语言差异、语言维持和转换、语言态度等方面的问题。下面先介绍语言差异的研究。

1. Liberson & Hansen (1974) 使用了 Greenberg 的公式 A 作为研究语言差异的工具，然后再运用了相关方法来观察使用的差异和国家的其他发展的关系。根据 77 个国家的数据，语言差异和 GNP 的相关为 −0.35，说明差异少和人均 GNP 高略有关系。另外两个指标则相关性略高一些，一个是城市化（r＝−0.52），一个是文盲（r＝0.49）。但是进一步纵向的研究表明语言差异和这些相关较高的指标不一定存在着因果关系。

2. Kuo (1979) 使用交际指数来表明新加坡和西马来西亚的社会语言型式。这两个国家的语言差异牵涉到 6 种语言：马来语、英语、泰米尔语、国语（即普通话）、福建话和广州话。下表为 1978 年调查的这 6 种语言的交际指数：

这两个社会的语言型式很不相同：除泰米尔语外，其他语言在新加坡的交际指数的水平大致差不多，所以新加坡人的交际靠的是学习其他民族的语言，而不是靠学习一种共同语。西马来西亚的情况则不同，马来语的交际指数最高，说明它已被选择为第二语言，起到是民族语言的作用。新加坡承认马来语为民族语言，但它还承认英语、国语、泰米尔语为“官方语

言"。福建话的交际指数是最高的，但它的跨民族的指数却很低，指数高是因为说福建话的社区特别大。

表 8.20　　6 种语言在新加坡和西马来西亚的交际指数

	马来语	英语	泰米尔语	国语	福建话	广州话
新加坡	0.453	0.381	0.408	0.004	0.607	0.399
西马来西亚	0.746	0.081	0.060	0.013	0.084	0.066

3. 双语补偿。公式 A 的问题在于它没有考虑多语制，这在公式 H 中得到纠正。Weinreich (1957) 提出另一个办法，测量双语制的补偿语言差异的程度。这就是说，如果差异高，双语制也会高一些。在那些说不同语言的人比较多的地区，人们就会多学会说另一种语言。他提出的公式是：

$$k = B/D \quad (8.11)$$

B 是人口中说双语的人的比例，D 是语言差异。Weinreich 发现在很多地区里的 k 都接近 1，说明语言差异为双语制所补偿。k 值低于 1 的地方多为山区和沙漠地带，那是交际较困难的地方。用公式 A 的符号也可建立一个测量双语制补偿的公式。A 指数为抽取两个不共享母语的人的概率。B 是人口中抽取一个说双语的人的概率。要保证两个人能够交际，只需要其中一个人是双语的。所以如果抽取一个懂双语的人的概率和从一对说不同语言的人抽取到一个说双语的人概率相等，差异就得到补偿。我们所需要的是双语制的比例除以公式 A 的一半的比率。

$$C = B/(A/2) \quad (8.12)$$

或

$$C = 2B/A \quad (8.13)$$

C 为补偿指数，B 为双语比例，A 为公式 A 的指数。

下 篇

实 验 方 法 篇

9. 语言学研究的实验方法

9.1. 定量方法

在第七章讨论定性方法时，我们也附带谈到定量方法，以兹比较。在第八章讨论话语分析时，在第九章讨论社会语言学研究方法时，我们也接触到定量方法。这说明这两种研究手段虽然很不相同，但从研究的课题着眼，它们并非互相矛盾的，而是互相补充的，可从不同的角度来加强观察问题的深度。在这里我们再集中介绍定量方法以及由此而派生的实验方法。

定量方法也就是计量方法，它和人类社会一起产生。人类生活的每一个方面都离不开计量：我们所生活的时间和空间，我们所生产的工农业产品，我们所吸收和消耗的物质的、精神的食粮，无不需要计量。计量方法也就是数学的方法，湖南教育出版社出版了一套《数学·我们·数学》丛书，从哲学、社会、经济、文化、教育、军事、语言、思维、创造等方面论述它们和数学的关系，可见计量的方法是我们须臾不可离的。《孙子算经》中说，“夫算者，万物之祖宗也。”世界文明史表明，所有的古代文明都是以发达的数学为标志的。古希腊科学的毕达哥拉斯—柏拉图传统就是相信世界具有数学表述的形式，科学的本质就是数学。我国古代教育中的“六艺”明确规定数学的教育。现代科学的发展更离不开数学，诸如力学、天文学、物理学、化学那些精确科学，在发展自己的理论时广泛地运用了数学的工具，以一些公式来表达自己的定律。在社会科学和人文科学中，也引入了数学的方法，例如：

在经济学中出现了所谓经济计量学（Econometrics）。1926 年 Frisch 提出这个词，1930 年成立了经济计量学会，1933 年发行《经济计量学杂志》。经济计量学是“经济理论、数学和统计学的结合”，其研究方法包括四个步骤：制定模型；估算参数；检验理论；建立模型。经济计量学运用了这些研究方法成功地进行市场需求分析、国民收入及其有关总量的计量分析、投入产出分析等等。经济计量学的主要课题是对各种经济量之间的关系提供分析方法，如因果关系分析、平衡关系分析、相关关系分析和时间序列分析。

在历史研究中也出现历史计量学，前苏联学者科瓦利琴科在其主编的《历史计量学》（1984）中指出，在科技革命的时代中，科学发展的一个明显特点是日益深刻的数学化和计算机化。数学方法在现代科学中日益广泛的普及首先是其发展内部需要所决定的。这些需要也充分表现在历史科学中。这本著作不但介绍了数学统计分析的基本方法，而且探讨了怎样运用这些方法来处理社会经济史大量史料和历史文献考证，并分析社会经济现象结构、社会经济过程动态、政治和文化现象，等等。

另一个使用计量方法的更为广阔的领域是和人有关的。为了对人（个人和集体）作出决策，我们需要教育和心理测量。在教育测量中考试占了重要的位置，作为考试的故乡，我国从隋唐开始就准备建立了科举考试制度。考试的根本目的是要鉴别人才和选拔人才，这就需

要把人的知识、能力和水平加以量化。人和其他的物质一样，由特殊物质（神经系统）构成的大脑及其技能（心理能力）是存在差异的。既然物质可以测定，心理能力也可以测量。这就进一步出现了心理测量学（Psychometrics）。心理测量把数学的方法引进心理学，借助一定的测量工具，对人的智力、能力倾向、成就、性格进行测量。

在语言学研究中使用定量方法，我们在上面已有所介绍。我们在讨论语料库建立时，已提到语料库语言学，其实它是计量语言学（quantitative linguistics）的一个分支。计量语言学涉及的范围要广阔得多，1994 年在俄罗斯莫斯大学召开的第二届世界计量语言学大会上决定成立国际计量语言学学会（IQLA，International Quantitative Linguistic Association），并出版自己的机关刊物《计量语言学杂志》（Journal of Quantitative Linguistics），杂志第一期的社论指出，计量语言学已经越来越多地把各方面的语言和语篇研究结合在一起，而且还要继续做下去。其中的一个原因是，计量语言学不仅是一门具有自己特定任务和目标的语言学分支，而且还具有更重要的特征：它是研究人员在考察语言和语篇现象、结构、功能和过程时所采取的一些方法和模型。另一个原因是使用计量方法和模型可以建立更新的、更丰富的概念，使人们能够对所观察的事物的各个因素之间的依存关系、相互关系和运作机制具有更深刻的洞察力。各种数学方法（概率论、随机过程、微分方程、模糊逻辑、集合论和功能论）正在被运用到观察语言和语篇的各种现象，包括心理语言学、社会语言学、方言学和语用学，也被运用到各个平面上的语言分析。在应用语言学中，诸如自然语言处理、机器翻译、语言教学、文件和信息检索等方面，人们对定量方法的兴趣也越来越高。

应该指出，社论仅是从计量语言学的角度，而没有从更广义的语言学的角度来谈论各种数学方法的应用，其实从认知语言学、心理语言学、应用语言学的实验性特征来看，计量方法还意味着实验方法和统计方法的结合，从实验的设计到数据的收集，分析和推断，都需要运用到数学和统计学的理论和手段。在这一部分里我们将着重介绍实验方法和统计方法。

在社会科学和人文科学的研究中引入数学的方法体现了当前科学发展中的文理渗透、学科渗透的特点，具有广阔的前景。在语言学研究中，一些边缘语言科学的产生更使数学的方法成为不可缺少的部分。

9.2. 定性方法和定量方法

从方法论的角度看，定量方法和定性方法（见 6.2）在几个主要方面都是不同的，故有人把它们看成是不容易调和的“两种文化”；但是这两种方法所得出的结果往往互相补充，有助于我们从不同的角度观察问题，故站在研究者的立场来说，不应有所偏颇。习惯于采取一种方法的人不妨从另一角度去看看，可以加深我们观察问题的洞察力。一般来说，从事社会科学和人文科学研究的人较熟悉定性方法，而对定量方法，特别是实验方法和统计方法比较生疏。

让我们首先看看它们的主要区别：

表 9.1　定性方法和定量方法比较

定性方法	定量方法
1. 自然观察	操纵和控制
2. 现象学观点："站在活动者本人的角度去了解人类行为。"	逻辑实证主义观点："对社会现象事实和原因的了解无需考虑个人的主观状态。"
3. 归纳	演绎
4. 综合	分析
5. 描述性	推断性

9.2.1. 操纵和控制

定性方法强调自然观察，包括直接观察、参与性观察和个案研究；定量方法强调对研究的环境的操纵和控制，即通过实验方法去观察。在实验的环境里，我们可以把一个要观察的问题的各种因素控制起来，而通过改变（操纵）某个因素，以了解因素之间的关系。这是实验设计的基本原则，后面还要详细讨论的。和自然观察相比，操纵和控制有它的优点和缺点，其实它们是互为补充的：

表 9.2　自然观察和定量方法比较

自然观察	操纵和控制
1. 观察面广，但分散	观察面窄，但集中
2. 变量不加控制，有利于了解它们的复杂关系，但容易顾此失彼	变量有所控制，有利于了解它们的因果关系，但容易简单化
3. 注意内容，但容易忽略形式	注意形式，但容易忽略内容
4. 解释力强，但容易主观	客观性强，但解释力弱
5. 接近现实，但时间长	时间短，但人为的成分大

9.2.2. 逻辑实证主义

逻辑实证主义是定量方法的哲学基础。在定量方法中我们必须通过推断来了解变量之间的因果关系。在外语教学研究中，我们感兴趣的是教学方法和教学效果之间的关系，但是影响教学效果有许多变量，如教师、学生、教学大纲、教学方法、教材、教具，等等，我们就首先控制其他变量，使它们相对地稳定，然后操纵要试验的教学方法这个变量，例如试验两种不同的方法 A 和 B，观察使用后的教学效果。由于其他变量都已经控制，我们就能看出方法和效果之间的因果关系。和逻辑实证主义相反的现象学则不同，它认为必须站在活动者的立场来了解社会现象，考察世界是怎样被经验的。经验的世界才是现实的世界，没有必要用人为的实验手段去了解世界。

表 9.3 逻辑实证主义观点和现象学观点比较

现象学观点	逻辑实证主义观点
1. 强调亲身参与活动以获得经验	强调用实验方法来获取数据
2. 只有通过个人主观经验才能认识人类行为	只有摆脱主观状态才能了解社会现象的因果关系
3. 了解（verstehen）就是移情	了解要保持距离
4. 依赖定性数据	依赖定量数据

9.2.3. 演绎

定性方法主要是采取归纳分析的方法，而定量的方法，特别是实验的方法主要是采取演绎的方法（见图 6.1）。在这类研究中，研究者一般（a）是根据观察所获得的印象（如自己的外语教学实践）；（b）或根据他人的某种理论（如某种外语教学模型）；（c）或根据别的研究领域所提出的，而又与自己的研究有关的理论（如认知心理学的关于场依赖和场独立（field dependence and field independence ）的理论对外语学习的影响），而形成假设。然后使用实验方法去验证假设，如果假设是有效的，就可进行推导和演绎。

表 9.4 归纳法和演绎法比较

归纳法	演绎法
1. 以数据为出发点	以假设为出发点
2. 没有事先形成的概念	进行预测
3. 可生成假设	检验假设
4. 研究成果：描述或假设	研究成果：理论

两种方法的区别可参看图 6.2。

9.2.4. 分析

定性方法主张采取综合法（见 6.2.4），综合就是把人们对于研究对象的各个部分的认识组合起来，形成一个整体的看法，其特点是从局部上升到整体，因此综合法也就是系统的方法。而定量方法则主张分析法，分析就是把研究对象的整体区分为各个部分，并逐一进行分析。自然界的事物都是复杂的统一体，它们都是由不同的部分组成的，其内在的因果机制往往不能直接感知，没有思维的分析活动，人们对事物只能有个混沌的感性的直观认识。分析法的特点是从整体深化到局部。综合和分析有不同的重点，但却是互相依存的。渡边茂等（1982）从系统的观点分析它们之间的关系：系统有外部结构，就是给该系统输入信号时，能得到什么样的输出信号？其着眼点是输入和输出的关系。当已给出内部结构时，把求其外部结构的工作叫分析。

综合则与此相反，已知的是外部结构（输入和输出），我们需要求其内部结构。

因此综合事物远比分析事物要难，为了综合，必须首先积累许多分析结果。美国在研制

阿波罗 11 号时使用了系统工程的技术，在阿波罗计划中能称得上要素技术的部分，据说多半采用分析法，而同时也采用较小规模系统（子系统）的综合法。

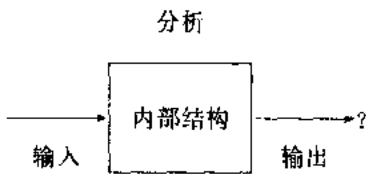


图 9.1

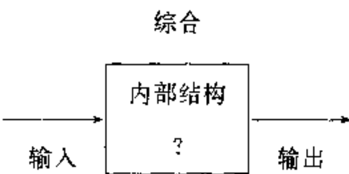


图 9.2

例如我们决定对学习者年龄这个成分和另一个成分——语言系统的语音子系统的关系，我们可以从综合/整体的角度去观察这两个成分的相互作用：即年龄对外语的语音系统的习得有些什么影响。但在这以前，我们最好采取分析法，例如专门观察某些元音的习得在是怎样受到母语语音系统的影响，既可以找一些评判员来进行主观的评判，也可以使用摄谱仪来分析这些音素。前者是定性的、综合的方法，而后者是定量的、分析的方法。由此可见这两种方法可以互为补充。

表 9.4　综合法和分析法比较

综合法	分析法
1. 从分体到整体	从整体到分体
2. 全面观	特殊观
3. 面向内部结构	面向外部结构
4. 了解过程	了解结果

9.2.5. 推断性

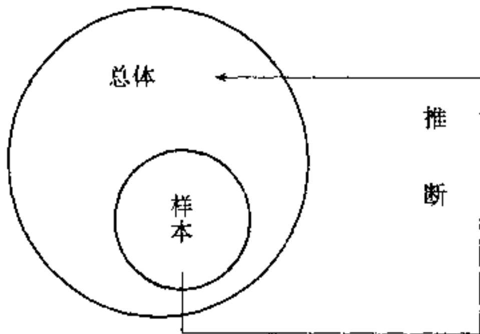


图 9.3　根据样本推断总体示意图

定性方法从自然观察出发，对所观察到的现象进行描述。我们在中篇的语言描写方法和社会语言学方法中已有较充分的论述。定量方法、特别是实验的方法，并不满足于简单的描述，它需要推断（inference）。这是因为实验只能在小样本中进行，但实验的目的却不仅是了解这个小样本的情况，而是需要根据小样本的情况来推断总体。要推断就必须使用许多推断统计学的理论和方法，例如抽样的原理、控制的原理、有效性的原理、无差别假设的原理，等等，我们在后面都要逐一讨论。

在这里我们仅是把描写性和推断性作一简单的对比：

表 9.5 描写性特点和推断性特点比较

描写性	推断性
1. 无控制的自然观察	有控制的实验
2. 归纳和描写数据	归纳数据进行推断
3. 旨在发现型式	旨在验证假设
4. 效度高、信度低	信度高、效度低
5. 概括程度低；个案研究	概括程度高：多元观察
6. 强调动态性	强调稳定性

我们在上面对定性方法和定量方法作比较，是为了看出它们的出发点、原则、手段的差异。这些差异正好说明我们在语言学研究中不能只限于某一种方法，而必须兼收并蓄。实际上它们可以放在一个量表上来对待：

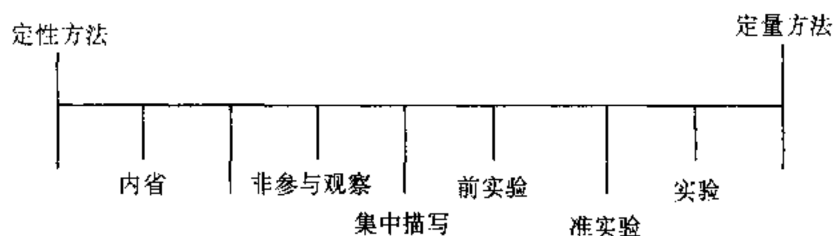


图 9.4 放在一个量表上的定性方法和定量方法

9.3. 实验方法的意义和应用

9.3.1. 实验方法的重要性

实验方法是科学研究经常使用的方法和程序。Kuhn（1970）认为，科学知识按照“前科学—常规科学—危机—科学革命—新的常规科学”的路线发展。常规科学（normal science）指的是“牢靠地建筑在一项或多项过去科学成果上的研究，这些科学成果都是某些科学社会所承认的，它们可以在一段时期内作为进一步实践的基础。……这是因为它们有两个主要的特征。一是它们是前所未有的，足以从相竞争的科学活动方式中吸引一群持久的支持者；二是它们是开放性的，足以提供各式各样的问题给不同组别的实践者去解决。我把具有这两个特征的科学成就称之为范式（paradigm）。”从前科学到常规科学的标志是范式的出现，在没有取得共同范式时，存在着各种各样的理论，需要发展和验证。就是建立了常规科学，也还会出现危机，导致新的常规科学的出现。Kuhn 的科学思想是科学的探索是一个不会终结的过程，我们需要不断提出新的科学假设，并进行验证。这就是 Kuhn 的历史主义的发现观。从这个角

度看，不但在常规科学中，就是在形成常规科学以前或以后，都必须进行科学实验。科学假设的本质是作出解释，而解释可以是多种多样的，Foss（1966）提出七种：

●因果性（causal），指某些直接的原因。

●历史性（historical），指过去的经验或事件。

●目的性（purposive），指某些意向或目的。

●规则性（rule following），指某些外部的规则，如社会规则。

●结构性（structural），指参与事件的有机体的成分。

●功能性（functional），指一般的，而非直接的原因。

●相依性（contingency），指某些平行的、有联系的事件，因而是相关的。

这些解释的相同点是它们均要求概括，也就是说，我们所找出的某种解释必须能够适用于相同的事件和个案。这里就牵涉到：

（a）怎样准确判断哪些事件和个案是相同的，这就是效度的问题。

（b）怎样把所得到的结果应用到更一般的场合，这就是信度的问题。

我们可以联系实验方法来回答这些问题：为了进行效度的检验，我们必须精心设计实验，保证参加实验的被试有代表性，在他们身上所取得的结果才不会是特殊的个案。在实验方法中，这是抽样的问题和描写统计学的问题。抽样保证被试的随机性，使相同的人都有同等机会成为被试；描写统计学整理结果，对样本进行描写，以明白地表示样本和总体的关系。描写统计学可帮助提高实验的效度。

为了进行信度的检验，我们必须用到推断统计学，其目的是了解我们的发现是否偶然的机会造成的。我们需要证明发现是有规律的，而不是偶然的。最简单的办法是把实验重复做一遍，但是这不很容易，因为实验条件会改变。在社会科学研究中常用的办法是使用推断统计学，它能告诉我们如果我们再做一次实验，我们在多大的程度上能够获得相同的结果。推断统计学关心的是结果的一致性，即信度。

应该看到，社会科学实验比物理科学实验要困难得多。这是因为它牵涉到人，而人是很难重复的，他们不像从口袋里捡豆子那样容易作随机抽样。实验室条件对所有的被试也不容易保持一致。实际上社会科学的实验并非在实验室做的，而是在现实的世界里做的。所以社会科学实验比物理科学实验的基础要薄弱。因为是在现实的世界里做的，有时就很难区别实验和实际的活动，例如让学生阅读一篇外语材料，就很难说这是在做实验，还是在训练学生的阅读能力。但是做阅读实验和训练阅读能力是不同的，做实验是为了取得实验数据，并根据数据来进行推断的。描写统计学和推断统计学正是把实验和实际的活动区别开来的重要手段。

另外，由于社会科学的实验牵涉到人，而人又是一个十分复杂的统一体。采取定性方法往往不容易把这些复杂的因素分离开来，逐一观察。以外语学习为例，我们常为学习者的成败参半而感到困惑。这是因为学习者的差异太大，有智力上的、生理上的、心理上的差异，而学习的外部环境又不一致。使用实验方法有助与我们把这些因素分开来，先从分体上细加考察，然后才形成整体的观点。按照 Tuckman（1972）的看法，实验方法有如下的特点：

1. 系统性（systematic）

实验方法是组织严密的系统，有一套必须遵守的程序性规则。这些规则包括怎样找出变量和定义变量，怎样设计实验以观察变量和决定变量的作用，怎样把数据和原来的研究课题、

假设联系起来等等。这些规则使实验系统化，便于操作和检查，同时也可以避免我们在考察中作出一些权宜的解决方案，克服由于研究者本人偏爱或其他外部因素对研究结果所造成的影响。和自然科学的实验一样，社会科学和人文科学的实验也倾向于精密，只不过因为研究对象是人，统计上的显著性意义应有别。例如在比较差异时，显著意义不一定要达到 0.01，达到 0.05 就可以接受。

2. 逻辑性（logical）

实验方法所执行的规则和程序形成一种直截了当的、逻辑性强的型式。它使研究一环接一环地展开，每一环和下一环都有逻辑关系，不可缺少。研究者可根据内部效度的要求去检查实验中的每一个环节，从而了解结论是否有效。他也可以使用逻辑来检查所概括的外部效度。研究效度的逻辑是我们决策的很有价值的工具，比我们靠观察数据，然后拍脑袋想出来的主意要高明得多。

3. 经验性（empirical）

实验方法从现实世界收集数据，这些数据以考试分数、学生在量表上的排列次序、学生达到某种水平的数目……等等方式表示出来。数据的类型虽多，但它们都可以量化的，也就是说，每一种数据都表示为一些可捉摸的数字。我们正是靠处理这些数字来和客观世界的研究联系起来。实验方法是一种经验性的研究，我们在上一部分所讨论到一些语言数据的描写的方法也属于经验性的研究。两者不同之处在于实验方法通过控制和操纵的方法，而描写方法通过自然观察的方法来收集数据。

4. 简约性（reductive）

当研究者对所收集的数据进行分析时，他把纷繁的个别事件和对象简约为可理解的概念和范畴。他牺牲了它们的一些特殊性和独特性，得到的是更为概括的事物间的相互关系，这实际上是对现实的内部规律进行抽象化的过程。简约化使实验不仅是在描写事物，而且是在解释事物。

5. 重复性和传递性（replicable and transmittable）

实验是可以重复的，这是因为实验的设计、数据的收集、统计和分析都有很大的透明度，别人可以重复这些过程来检验实验结论是否正确，别人也可以使用研究的结果来进行另一项研究。

9.3.2. 实验方法在语言学的应用

现代语言学有许多分支，这些分支都是在一些学科的边缘上产生的。如以语言学为主体，它和社会学结合就产生社会语言学，和数学结合就产生计量语言学或数理语言学，和心理学结合产生心理语言学或认知语言学。和教育学结合就产生应用语言学（狭义），这些结合不但拓宽了传统语言学的视野，同时也引进了很多相关学科的研究方法。

实验方法是应用语言学和心理语言学（包括认知心理语言学）最常使用的研究方法。神经语言学也使用了这个手段，但它还使用了很多医学的临床观察的手段。

9.3.2.1. 实验方法在应用语言学中的应用

一般来说，应用语言学有广义和狭义之分。广义应用语言学指的是联系实际问题研究语

言和语言学；狭义应用语言学指的是对第二语言与外语教学的研究（见 Richards，1985）。应用语言学是一门实验性很强的学科。无论是广义的或狭义的应用语言学，都并非“纯”理论的研究，而是有特定的“问题求解”的任务。（桂诗春 1988，1993）因此要采取较多的计量方法和统计方法。

应用语言学（狭义，下同）在语言学和教育学的边缘上产生，也有人称之为教育语言学。它吸收了教育实验设计和教育测量的理论和方法，发展成为一门很有生命力的独立的学科。应用语言学有基础理论的研究，如探讨外语教学的认知基础和过程，检验语言习得的理论，提出外语教学模型，考察学习者的差异等等；也有具体应用的研究，如教学大纲、教材、教学法、电化教学设备的制订和使用，教学评估和测试等等。应用语言学的研究范围十分广阔，而且很多因素又纠缠在一起，不易分开。Paivio & Begg（1981）谈到采取实验方法研究第二语言教学的必要性：“实验是对自然的有控制的观察。实验者建立一个这样的实验环境：研究任务的结构很明确，观察的行为的性质也很明确，所提出的问题也很具体确切。”

1. 实验方法是解决外语教学中有争议的问题的一个有效方法。例如在外语教学法中围绕着外语和母语的关系有过长期的争论，Stern（1983）认为这是外语教学中三大难题之一。这个争论体现在教学法上就是语法—翻译法和听说法之争。60 年代中叶在西方兴起一股建立语言实验室之风，当时有不少人以为语言实验室＋听说法是外语教学的出路。但是这股风在很大的程度上是那些语言实验室的厂家吹起来的，究竟是否那么神奇？美国的外语教育学家 Rivers 和心理学家 Ausubel 都指出这种方法和先进的教学理论是背道而驰的（见 Smith 1970，Keating 1963），Keating 首先在 5000 名学法语的学童中观察实验语言实验室的效果，发现所有显著性差异都是有利于不使用语言实验室的学童的。这更促使美国外语教育学家决心了解其真正的作用，于是在 Lado，Sapon，Staar，Twaddell 等一批专家指导下，美国宾州政府组织了一次规模巨大的实验：宾州外语教学实验（The Pennsylvania Foreign Language Project），以比较外语教学中认知法和听说法的优劣。试验从 1965 年开始，延续了四年，第一年参加试验的有 120 名教师，在四个地区的 58 个学校和班级里进行，教授的是法语（61 个班）和德语（43 个班）。试验对比研究了三种教学方法：传统法（认知法）、功能技巧（听说法）、功能技巧＋语法。在后两种方法中还使用了三种不同类型的语言实验室。其实验的组织如下：

传统法	X		
	教室＋录音机	听力主动型语言实验室	听力录音型语言实验室
功能技巧＋语法	X	X	X
功能技巧	X	X	X

图 9.5　宾州外语教学实验的组织方案

参加试验的教师都经过严格的挑选，以保证他们的教学水平和教学经验基本一致，然后他们还需参加一周的教学法训练班。教材是经过选择的。测量工具也是一致的，试验前统一

参加了加州心理成熟测试和现代语言学能测试；在试验中使用了现代语言协会的合作外语测试，包括有听力、阅读、写作等内容，另外还增加了一个口试。对学生还发了一个教师意见调查表。试验收集了大量的数据，这里无法一一介绍。下面仅引用其中几个重要的结论：

（1）三种教学方法在培养听、说、读、写四种技能方面的比较：

听：三种方法没有显著性差异。

说：三种方法没有显著性差异。

读：传统法比功能技巧方法好，有显著差异；传统法和功能技巧法＋语法没有显著差异。

写：三种方法没有显著差异。

（2）在三种语言实验室系统中，哪一种对训练发音和结构准确性在经济和教学方面更为合适？

在第一阶段，(a) 使用录音机和每周在（b）听力主动型或 (c) 主动录音型 语言实验室增加两次练习的效果并无显著差异；在第二阶段结束，这三组亦未见显著性差异。语言实验室对听和说未见明显效果，但在语言实验室所花的时间可能影响阅读能力。

（3）有哪些变量或变量组合可以预测学生的外语成绩？

除了哪些微小的差异外，在学能测试中的英语（即母语）和社会科学的分数最能预测外语学习的成绩，对不同的语言技能来说，其多元相关系数在 0.51 到 0.89 之间。

（4）比较学生对这三种方法和三种语言实验室的态度。

学生从开始学习到最后结束的外语学习态度是越来越低的，但是这种下滑和教学方法和语言实验室无甚关系。对 215 名学生的单独访问表明他们对实验的态度无显著差异。

另外还有一些对教材和教师的调查，亦无大差异。在第三年和第四年，也继续作了一些观察，因为有些班级的学生退学较多，较难比较。这里不再列举其结果。与此同时英国的 York 大学在 Hawkins 教授的主持下也作了类似的实验，结论亦大同小异（见 Green，1975）。由此看来，实验方法有助于我们拨开云雾，看清事物的真相。

2. 另一个值得一提的例子是学习外语年龄的问题。长期以来，人们凭一些零碎的观察，得出外语学习要从越小学起的印象，于是在我国一哄而起，提早在小学教外语，走入了一个教学误区（桂诗春 1992）Cook（1986）专门以外语学习的年龄因素为例说明在第二语言教学研究中应用实验方法的重要性。她列举了 22 项实验研究，然后归纳出这样的结论：

（1）在第二语言学习时，年龄大的儿童要比年龄小的儿童学得好些。在 22 项报告中，有 21 项说明年龄大的儿童要学得好些，如一年级的比幼儿园的、14 岁的比 11 岁和 8 岁的、15 岁的比 5 岁的、17 岁比 8 岁的、15 岁比 8 岁的、9 岁比 4 岁的、9 年级比 1 年级的。而且这还包括语音的学习。

（2）在第二语言学习时，成年人要比儿童学得好些。有的报告表明，成人比所有年龄组都要好些。有的研究指出，“在外语发音方面，成人比儿童高明。”美国著名教育学家 Thorndike 相信，20—40 岁的成年人比 8 岁、10 岁或 12 岁儿童要学得好些。有的研究认为，30 岁是外语学习的最佳年龄。不过这些结论大多是根据参加短期训练的，特别是在学习者自己的国家学习的结果而作出的。

（3）早一点学习第二语言的移民要比迟一点学习的移民要学得好些。很多研究使用了回顾的技巧来观察成人学习第二语言的情况，结果都支持这个结论。有的研究指出儿童移民口

音的多寡取决于其到达移民地的早晚，到达越早的第二语言中的母语的口音越小．但是有的关于英国人移民到荷兰的研究都得到相反的、复杂的结果。

这三个结论看似有点矛盾，其实是很好解释。第一、第二点和第三点不同，主要在于学习者是在自己的国家还是在使用目标语的国家学习第二语言。移民的某些特点，不管是在学习者自己身上还是在环境方面，都足以使成人产生自卑心理，因而影响其学习。

应该指出，正如定性方法和定量方法是互为补充一样，外语教学的研究虽然较多地采取了实验方法，它并不排除使用其他方法，例如自然观察。有些语言习得的研究使用了长达数年的纵向观察，德国 Preyer 在 19 世纪末期逐日记录了他的儿子头 3 年的语言发展；Stern 夫妇 1907 年发表了他们对三个儿童学话的详尽实录。Leopold 连续观察了他的女儿的双语（英语和德语）发展达 10 年（1939—1949）之久，整理成 4 卷资料。我国赵元任也观察了他的 28 个月孙女的汉语学习过程。在第二语言学习方面，人们又把注意力转移到课堂教学的研究，在这些研究里也采用了很多观察和记录学生的言语行为的手段。最近香港中文大学李行德等还建立了一个香港中国儿童语言习得的语料库，供研究者使用．在国际互联网中可以检阅和调用。Seliger & Long（1983）指出，为了要很好地记录教师和学生在课堂上的交往，外语教学研究者发展了近二十种记录的系统，观察的重心也不同，有言语的，有辅助语言的，有非语言的，有认知的，有感情的、有教学的，有语言结构的，有语篇的。

9.3.2.2 实验方法在心理语言学中的应用

心理语言学使用了很多心理测量的方法，用以揭示语言和认知、语言和心理的关系。不管是语言的学习还是使用都是一个信息处理的过程，这个过程牵涉到和外部环境有密切关系的输入和输出，也牵涉到处理信息的人脑。人脑就相当于计算机的中央处理器，在这个处理器里活动的是包括知觉、表象、记忆、理解、意识、思维、决策、问题求解在内的各种心理表征。（见桂诗春，1992）心理语言学有基础研究，如语言和记忆、语言的理解和产生、语言和认知、语言和思维的关系究竟是怎样的？心理语言学也有应用研究，如语言习得过程、语言学习者的内部大纲、语言学习者的个别差异等等（见王初明，1990）。语言是可观察的，但心理活动却是看不见、摸不着的。所以我们往往通过语言来了解心理过程，难怪有人说“语言是心灵之窗。”但是语言也难以控制，靠自然观察的方法往往费时失事，这就只好采取实验方法。在心理语言学研究的早期(60 年代)，大多数实验都是用来验证语言模型的心理现实性，例如关于 Chomsky 的表层结构和深层结构的实验。但是到了 70 年代，心理语言学家觉得要独辟蹊径，于是出现所谓“没有语言学的心理语言学”时期，“结果是，心理语言学家割断了他们和语言学的联系，被吸引进认知心理学的主流里面。”（Tanenhaus，1988）对这以后的蓬勃发展，读者可参阅 Tanenhaus 的概述，这里我们不妨看几个例子：

1. 人的短时记忆容量有限，而且保留信息的时间很短。所以我们在短时记忆里尽量把有关的意念组合在一起。Bransford & Frank（1971）的实验就是企图验证这个设想的：他们设计了 4 个复杂句，每一个句子有 4 个“意念”，从意义上看，就是命题。每一个命题都可以和其他的一个或多个命题组合而成为更复杂的句子，如其中一个复杂句是：

The ants in the kitchen ate sweet jelly which was on the table. [厨房里的蚂蚁吃了桌上的甜冻糕。]

句子由 4 个命题组成：

a. The ants were in the kitchen. [蚂蚁在厨房里。]

b. The ants ate the jelly. [蚂蚁吃冻糕。]

c. The jelly was sweet. [冻糕是甜的。]

d. The jelly was on the table. [冻糕放在桌上。]

然后他们把这些命题加以组合，构成有一个、两个或三个命题（但没有全部四个命题）的句子，如：

The ants in the kitchen ate the jelly. [在厨房里的蚂蚁吃了冻糕。]

他们然后选了 24 个句子，把这些组合的句子混在其中，读给被试听，让他们作“是/否”试验，判断哪些句子是原来听到过的句子，而且还对自己的反应的信心打分。结果是只要句子和原来句子的信息相一致的，被试就无法说出哪些句子是听过的，哪些是没有听过的。而且越是复杂的（即命题多的）句子，被试就越有信心说他听过这个句子，如 The ants in the kitchen ate the sweet jelly which was on the table.，其实他并没有听过这句话。Bransford & Franks 由此推论，被试注意的不是句子的表层的句法，而是把几个句子组合在一起的单一的意义表征。他们在回述（recall）时，把能够紧密衔接在一起的哪些句子认为是原来听到的句子。这就是说人们爱把意义组织成一个单一的表征放在记忆里。

2. 汉语使用一种和西方拼音文字很不相同的象形文字，研究汉语使用者的心理语言机制是一个越来越引起心理语言学家的兴趣的领域（见 Kao，H. & Hoosain，R.，1984）。例如有些研究发现，在辨认单个汉字时，使用左视场较为方便，而右视场在辨认两个字组成的词占优势，因此大脑两半球在辨认汉字时似有不同的功能。Hatta（1978）企图从视觉处理和语音处理的角度去解释这个问题：当被试看到“生”时，因为缺乏语境，没有语音形式，所以使用管辖形象思维的右半球处理；但是如果看到的是“先生”，它就具有语境，有语音的形式，于是就使用管辖语言的左半球来处理。曾志朗等（1979）则从右半球的整体感知和左半球的分析/序列功能的区别来解释。单个汉字是整体感知的，而两个字组成的词则需要分析/序列处理。但是怎样进行具体的分析/序列处理，仍然不很清楚。何世强（译音）与 Hoosain（1984）于是作了一个两半球在对立面感知中的区别的研究。他们让被试看三种不同类型的搭配：一种是真正的对立面，如“问答”、“南北”、“美丑”；第二种是一个为同音词的非对立，如“正付”（“付”为“负”的同音词，从声音上是对立的，从文字上是非对立的）、“左又”（“又”为“右”的同音词）、“始中”（“中”为“终”的同音词）；第三种为真正的非对立，如“善很”、“深夫”、“好电”。实验使用特制的速示器（tachistoscope），将这些“词组”显示在注视点的左或右边，然后要求他们作出判断。结果表明左/右视场的主要效应没有显著意义，但是正/误的主要效应有显著意义。视场和对立组/非对立组的交互作用有显著意义。特别是右视场处理对立组要显著地快。这意味着对立组的信息主要是传递到左半球。非对立组是和对立组刚刚相反，但两半球的处理时间却没有显著差别。实验者认为，这可能是因为被试需要更多的反应时间来进行反复检查。换句话说，如果大脑的认知功能是有所侧向（lateralization）的话，必须能够充分地进行一次性的直接处理，才能取得效果。如果直接处理不足，就必须调用大脑其他区域的资源。

3. 心理语言学的另一个热门的研究课题是心理词汇。一般认为词和意义是用网络的方式保存在大脑里，在语义网络里有许多节点可以激活和扩散，以引起联想。对词和意义的关系有不同的解释：一种解释认为词和意义是保存在一起的，这可以说是直接提取模型；另一种

解释是词和意义是分别保存的，这可以说是检索模型（桂诗春，1991）。这些解释对外语教学有不同的含义：按前一种说法，新学的外语词汇和意义一起保存，而母语的词汇又是和意义分别保存的；按后一种说法，只有一个统一的心理词汇，但是两种语言的词形和声音的表征保存在提取档里，然后根据和输入匹配的表征，再到心理词汇中提取意义。和这个问题有联系的是词汇提取时是否要通过语音的重新编码（recoding）。英语是拼音文字，文字和声音的关系很密切；但是汉语是象形文字，文字和声音的关系没有那么密切。那么中国学生学英语又会出现什么问题呢？桂诗春（1992）的视觉词汇辨认就是企图回答这方面的问题的。他让被试看一对对的词来作判断，一些是真正的词（如“邮局”和“strong”），一些是不存在的非词（如“概睬”和“mtuis”），一些是伪音词，它们并不存在，但从语音规则来看，是可能的（如“基会”和“brane”）。这一对对词有6种搭配：有的意义上有联系（如父亲—儿子，stamp—邮局，strong—weak）；有的声音上有联系，而意义上没有联系（如单价—担架，kites—开始，due—dew）；有的两者均无联系（护士—理由，孩子—matter，last—optical）；有的两个都是非词，英语是伪同音词，汉语是声音上和另一个词相近（刚笔—丙干，brane—基会，neskt—leiber）；有的两个都是非词，英语为真正的非词，汉语是声音上和无甚联系的非词（太亩—发交，cikra—池余，treisp-bseer）；有的两个词，一个是真正的词，另一个是非词，英语为伪同音词，汉语为声音上与另一个词相近的非词（符号—沉年，strain—柠檬，hut-gat）。实验让被试（学英语的研究生）在计算机屏幕前按键作出反应，并把反应时间（毫秒）记录下来，以作比较。结果说明（1）对这些学了比较长时期的英语的人来说，两种语言只有一个统一的心理词汇，激活扩散可以在两种语言中交叉进行。但被试对两种语言的掌握程度不一样，汉语的反应时间最快，英语较慢。如果两种语言交叉，则反应时间缩小，英语和汉语无显著意义的差别。（2）在视觉上，英语的语音表征促进英语的词汇检索，而汉语的语音表征不会促进，说明被试在视觉上处理汉语时，并没有进行语音编码。（3）中国的英语学习者经过一段时期的英语学习后，能掌握英语的语音规律，他们拒绝英语伪同音词所需的时间多于真正非词所需的时间。而在汉语和英语交叉的情况下，却没有这种现象。（4）英语和汉语属不同的文字系统，词汇提取在语码转换时会出现干扰，但如果两个词属于一个语义网络，也会出现相反的促进作用。

9.4. 认知科学与实验方法

在介绍应用语言学和心理语言学的具体实验设计过程以前，有必要讨论实验方法的一些新发展，因为现代科学的发展使我们的思想和研究手段都大大更新了。Gagne，E.（1985）认为当前认知心理学研究方法上的进步相当于50年代遗传学研究中的突破。在50年代以前，遗传学家已经认识到基因可以使有机体的某些特征可以按一定的预测的比例遗传给后代，但是基因究竟是如何发生作用的，却没有什么实验性的数据去说明。接着，研究基因的手段出现了重大的突破，X光结晶学的完善使遗传学家发现了遗传物质DNA的结构，蛋白质合成和放射性追踪技术又使我们有可能破译遗传密码。正是利用了这些技术，科学家才能清楚基因是DNA的一部分。这说明研究手段的发现有时候能够促使研究的发现。

当前我们面临着的形势是认知科学的崛起，它把各门有关的学科组合在一起，成为一股

巨大的力量，势如暴风骤雨，不可阻挡。认知科学从计算的角度研究智能和智能系统。图 9.6 显示了一些学科对认知科学的建立所起了重大作用。

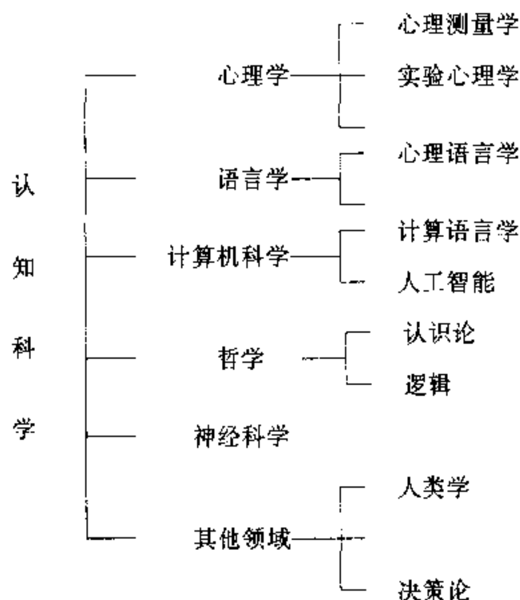


图 9.6 认知科学的组成部分

9.4.1. 认知科学的两条基本原则

认知科学认为人类认知包括可计算的机制（computational device）和表征的机制（representational device）。Eckardt（1993）指出，"认知能力包括一个具有计算能力和表征（信息处理）能力的系统。"由此产生两条基本原则：可计算性的原则和信息处理的原则。这两条原则奠定了实验方法的基础。

9.4.1.1. 可计算性的原则

这个原则似乎不大好理解，人们的心理过程和认知过程复杂多变，往往连我们自己都说不清楚，怎么能够用计算方法来表示呢？要回答这个问题，认知科学家用了一个所谓"计算机的隐喻"，把人脑比作电脑。一般人以为计算机是用来算数的，其实这是误解。如果光是算数，计算机的用途就很有限，它只能处理数目，不能处理文字和图象，更不能具有智能。其实计算机处理的是符号（数目字），这些符号可以用来表示五光十色的世界，也可以用来表示我们的心理过程中的知觉、意念、表象、假设、思维、记忆等（在心理学中，称为心理表征）。计算机有两种符号功能：一是操纵符号，既可以改变符号，也可以创造新符号；二是它可以把操纵符号的指令作为程序保存下来，并逐一执行。但是计算机不能把符号和客观世界联系起来，因此它运算得来的结果必须由人去解释。人脑和计算机相似，它也是一个符号系统，它在认知过程中可以操纵符号，可以创造新符号，而且还可以把得到的结果和客观世界联系起来，作出解释，例如验证心中的表象和客观的描述是否相符。而且我们心中的符号不一定和客观事物相对应，它可以是一些假设（如果我是一个学校的校长，我就……；假定这

个地区发生地震，就应该……），也可以是一些虚拟的表象（如龙的传人，钟馗抓鬼，八仙过

是这些表征是怎样作为表征而起作用的，却仍然不很清楚。Eckardt 认为早在认知科学以前，哲学家 Peirce 就提出了表征的一般理论（即他的符号学理论）。Peirce 认为符号（sign，他常称之为 representamen），对象（object）和解释者（interpretant）存在着一种三角关系。

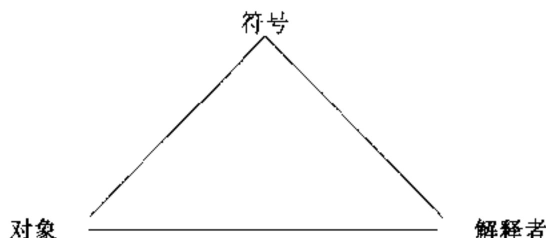


图 9.7 符号、对象、解释者的三角关系

对象是符号所指的可感知的人类经验，而解释者则是对符号的理解者。符号在对象和解释者之间起传递的作用。符号之所以成为符号是因为它代表了对象和解释者之间的某种关系。符号和它所表示的事物并不完全相同，它有其自身的特性，和它的表征功能不一样。Peirce 称之为“物质性能”。例如在浴室里的一对水龙头，其中一个带有 H 的符号，它的物质性能就是两条直线加一条横线。但是它表示的是热水的水龙头，这是解释者对符号的理解。而符号所指的对象则是一个扭开后热水可源源而出的开关，这是可感知的人类经验的一部分。人们对热水龙头的心理认识和取得热水的实际经验是靠 H 这个符号联系起来的，而 H 本身与对象（水龙头流出热水）和解释者（热龙头）并无必然的联系，只有放在三角里才有意义。那么怎样从 Peirce 的框架来看表征呢？Eckardt 是这样归纳的：一个表征机制具有表征状态或在机制里载有表征的客体，这些表征的总体可以称为心理表征系统。任何表征必须具有四个主要特征才能成为表征：（1）它必须由一个表征载体（符号）来实现；（2）它必须代表一个或更多的表征对象；（3）它的表征关系是多少有根据的；（4）它必须是某些目前存在的解释者能够解释的。

由此看来，心理表征其实是客观世界和主观世界联系的纽带。客观世界作用于主观世界才能产生种种心理表征，这种作用是有序的，所以我们又可以把人类看成是一个有限度容量的信息处理系统，大脑在执行某项任务时所使用的机制和过程可以理解为一件任务被分解为若干阶段，然后作有序处理的信息流程。例如一个外地来客向你打听，“请问五号无轨车站在哪里？”你必须听懂每个词，检索出其意义，确定这些词组合在一起表示什么，了解到对方所提出的问题，再到心里找寻答案，定出一个作出回答的计划，然后把计划变成真正的答案：“往前面走 100 米，再向左拐。”我们可以追踪这个过程的内心信息流。信息表示我们所操作的各种心理表征：问题、它所表示的意思、车站在哪里的记忆、制定回答的计划等等。这些心理操作的过程具有有序化的特征。信息处理分析的最重要的特点是，它必须对有序的心理过程及其所产生的结果进行追踪或回溯。这就产生信息处理的原则。信息处理分析常表现为流程图，一个流程图有若干框架，每一个框架表示一个处理阶段。箭号表示从一个阶段到另一个阶段的暂时的过渡。信息处理过程可以表示为下面的流程图（Gagne，1977）：

在这个模型里，信息以能量的形式出现（视觉中的光、听觉中的声音、触觉中的压力等等），接受能量的是对它特别敏感的接受器（如耳朵和眼睛），接受器以电化冲动的形式，将

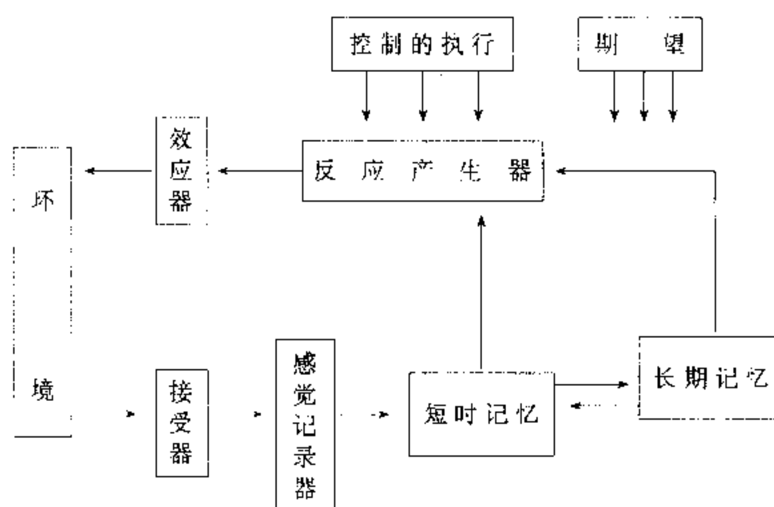


图 9.8 信息处理的流程图

信号传到大脑。这些神经冲动首先到了神经中枢的感觉记录器[①]。每一种感觉都有其不同的感觉记录器，但是它们都是感觉信息在神经中枢的真实表征。保存在这些记录器的时间很短，约1/4秒。在全部感觉信息中，被保存下来的只有一小部分，以便在短时记忆里进行处理。这个信息减少的过程可称为选择性知觉的过程[②]。短时记忆就是对事物的意识（awareness）。当我们在一段时期内意识到某种事物的存在，就可以说该事物保存在短时记忆里。短时记忆的容量不大，约为 7±2 个单位[③]；时间也不长，约十秒钟。但是单位可大可小，可以是孤立没有联系的单位，也可以是意义上有联系的较大的单位。所以人们可以通过联想来扩大记忆容量。在短时记忆里的信息可以进行编码，然后存到长期记忆里。长期记忆把信息储存起来供以后使用，有的信息可以保存很久，甚至一辈子。信息检索是产生反应的基础（如在上例中先到记忆里寻找汽车站的地址，再回答问题），在有意识的思维活动中，信息从长期记忆流向短时记忆，然后抵达反应产生器。在自动化的反应中，信息在检索过程中直接从长期记忆中流向反应产生器。反应产生器把反应的程序组织好，并指导效应器执行程序。效应器包括我们所有的肌肉和腺。信息流是有目的、有组织的，受到两方面的影响：一为执行的控制过程，一为期望。这两个过程是个人在信息处理前学到的经验。它们可以说是长期记忆的另一部分，其功能是决定（或选择）完成某种任务所需的信息处理程序，例如决定采取哪一种或哪几种处理信息的方法——注意什么信息、储存什么信息、对哪些信息进行编码和检索等等。期望和我们完成任务的动机有关，我们想完成什么可以影响我们注意什么，对什么信息进行编码。例如我们期望的是了解电路版的电阻情况，我们就不会去注意电阻器的形状或电路版在仪器中的位置。如果我们在交际中期望的是信息的内容，我们往往就不会去注意信息的形式。

许多心理语言学的实验中都采取了信息处理分析的方法，下面试举一例。Clark & Chase 作了一个关于句子理解的实验，让被试判断所看到的句子和所显示的图形是否一致：

① sensory register，亦可称为 iconic memory。

② 知觉是有选择性的，这就是"注意"，未被信息接受人注意到的信息就会很快地丧失。

③ 这是 Miller（1967）的估计，见他的著名的论文：The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information，载于 The Psychology of Communication（Middlesex: Penguin）。

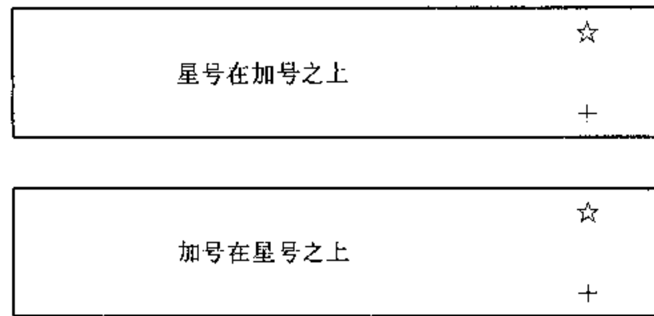


图 9.9　被试在句子理解中见到的两个刺激

Clark & Chase 预测被试在判断句子和图形不相符时所花的时间要比两者相符的时间要短。实验的结果证实了他们的想法。但是怎样解释这种现象呢？他们认为这是因为两者的信息处理过程是不同的，图 9.10 是一个句子验证的信息处理分析流程图：

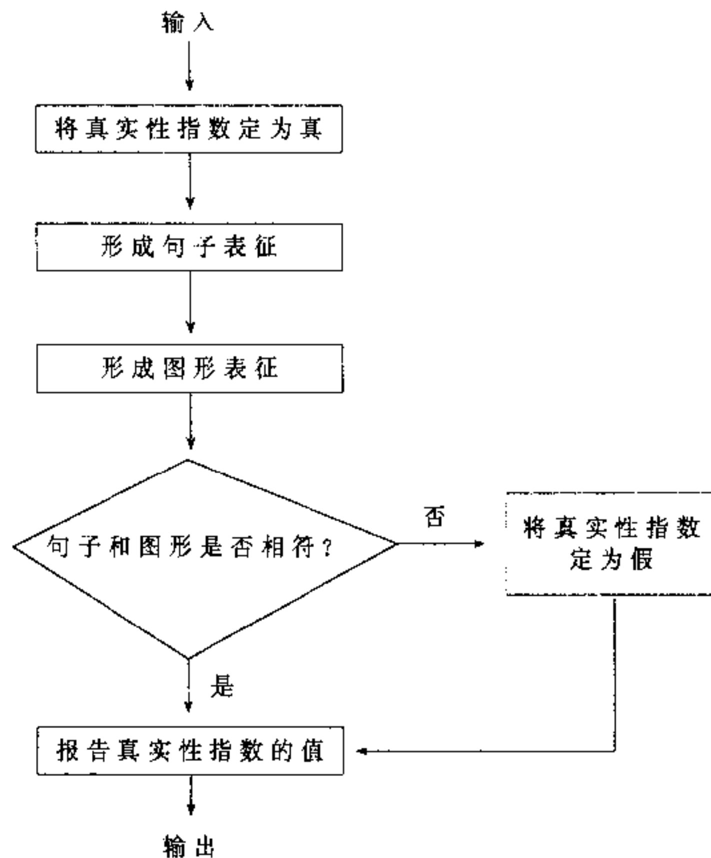


图 9.10　句子验证的信息处理分析流程图

Clark & Chase 的假设是被试在判断句子时首先把真实性指数定为真，然后建立句子和图形的表征加以比较，如果两者相符，就报告这个真实性指数的值。如果不一致，就要将真实性指数改为假，然后再报告。所以如果不相符，就需要多一次操作。故需要多一点时间。

应该指出的是，流程图可以提供的仅是一种解释，而不是唯一的解释。对同样的现象，Carpenter & Just 提出另一种流程图，这就是所谓对实验结果的解释的不确切性（non-identifiability）问题，可参阅桂诗春（1991）。

9.4.2. 一般的实验方法

9.4.2.1. 表征类型分析法

认知科学所研究的典型课题是弄清楚某一信息在内心中表现为什么形式。例如某一个离散的刺激（词或字母）在它显示后究竟是以什么样的形式保存在记忆里的？这个形式又怎样随着时间而发生变化？表征类型分析假定一个刺激可以分解为一系列的编码阶段，每一个阶段都提供一个刺激的临时的"内部记录"，供别的过程阅读和利用。这个内部记录就是表征(representation)，在每一个刺激分析阶段的表征可成为"内码"，我们研究的目的就是描写这些内码和它们的性质。

这种理论的基本看法是，一个离散的刺激事件产生某一种模态（视觉的、声音的等等）中的知觉性的内码，然后视指令的不同，这些起始的编码又会引起一些第二位或第三位的内码。例如一个字母可以开始时表示为一个视觉的形式，而这种形式又可以触发起一个音素码，乃至触发起一个辅音或元音。几个内码可以同时共存，但是有些内码可能在刺激分析的早期阶段就能提取。人们感兴趣的是在各种内码中区别出哪些内码是已经形成的，而且是在什么时候形成的。

解决这个问题的方法有多种，基本的方法是让被试快速辨认两个刺激（字母、句子、图画……）在某一状态下是否相同。例如让被试决定两个连续出现的形式是否外型相同（AA 为相同，Aa 和 Ba 为不同），或是否名字相同（ Aa 为相同，AB 为不同），有的研究者发现在名字是否相同的试验里，辨认两个既是名字相同，又是外型相同的刺激，要比只是名字相同的刺激快 70 到 100 毫秒。因而得到这样的结论：被试在形成内码时保存了其外部特征，而这个外部特征又被用来进行第一个和第二个字母的外形匹配。桂诗春和李崴（1992）使用回述的方法，发现学习英语的中国学生在处理英语时经过语音编码，而处理汉语时却无须经过这个过程，也是一个表征类型分析的实例。

9.4.2.2. 扣除法

早在上一世纪 80 年代，Donders（1969）就提出了扣除法（subtractive technique），这可以说是一种直接测量心理过程的方法。这种方法使用潜伏性数据① 作为数据源，其基本出发点是认知过程是可以分离的，一个反应是一系列线性操作过程的结果，每一次操作都需要一定时间，而这些时间可以递加起来。使用扣除法的一个典型实验要求被试作两件略有不同的作业（如作业 1 和作业 2），其中一件作业牵涉到多一次认知操作（T_3）；而其他的认知操作对两件作业都是一样的（均为 T_1+T_2）。完成每一件作业的时间都记录下来，然后将完成一件作业的时间从另一件作业的时间中扣除出来。所剩下的时间便是完成那个多出的认知操作的时间。

作业 1 的时间$=T_1+T_2+T_3$

作业 2 的时间$=T_1+T_2$

---

① latency data 是实验中所收集的关于反应时间的数据，故也称为反应时。

T3 的时间＝作业 1 的时间－作业 2 的时间

Clark & Chase 用信息处理过程的实验方法所做的句子和图形的判断实验其实也是采用了扣除法，读者可参看图 9.11。

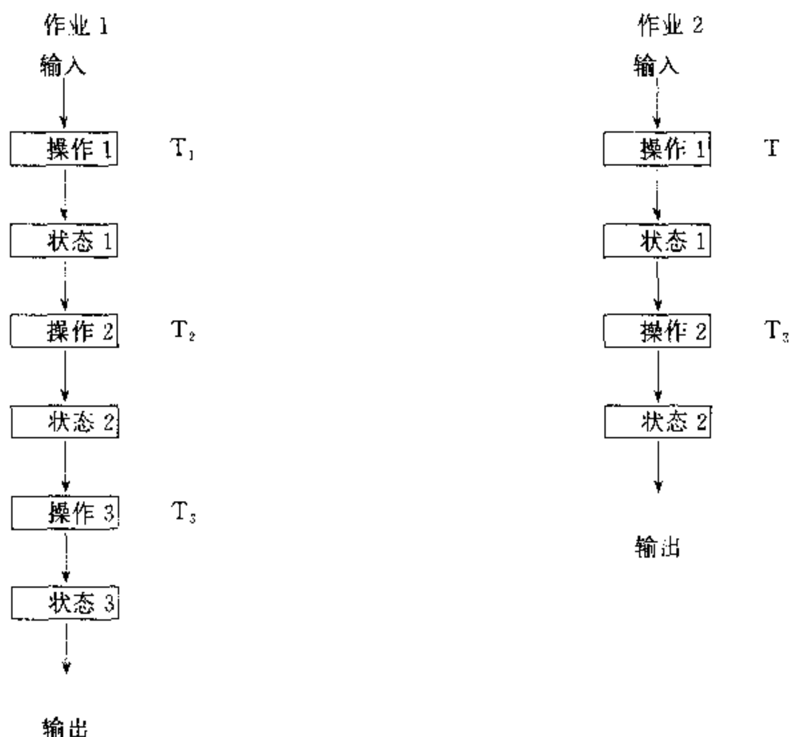


图 9.11　扣除法示意图

9.4.2.3. 递加因素法

递加因素法(additive factor method)是 Sternberg (1969) 在扣除法的基础上发展起来的，他认为，完成一件作业的总时间是每一个处理阶段的时间的总和，如果我们使用一些方法来操纵每个阶段的处理过程，就可以了解其处理时间。所以这种方法无需减少一个处理阶段。这种方法强调处理阶段的递加性和独立性，其基本出发点是，如果两个阶段是独立而又有连续性的，我们就有可能找到一个只影响一个阶段，而有不影响另一个阶段的试验变量。如果我们对反应时的原始数据做方差分析，就会发现因素之间没有相互作用。使用这种方法有几个前提：一是要完成的作业必须包括一系列独立的处理过程；二是在每一个处理阶段，从上一处理阶段获得的信息经过某种转换后会传递到下一个阶段；三是在任何阶段的信息转换的性质和产生它的速度都不受前面处理阶段的时间的影响。后一点是争议较多的。（见 Eysenck，1988）

Sternberg (1969) 所设计的高速记忆扫描试验是一个使用递加因素法的例子。试验向被试显示一些记忆项目的集合，然后再显示一个试探词。如果试探词是显示集合中的一个，被试就要作出“是”的反应；否则就要作出“非”的反应。他把被试对试探词作出反应的时间记录下来。

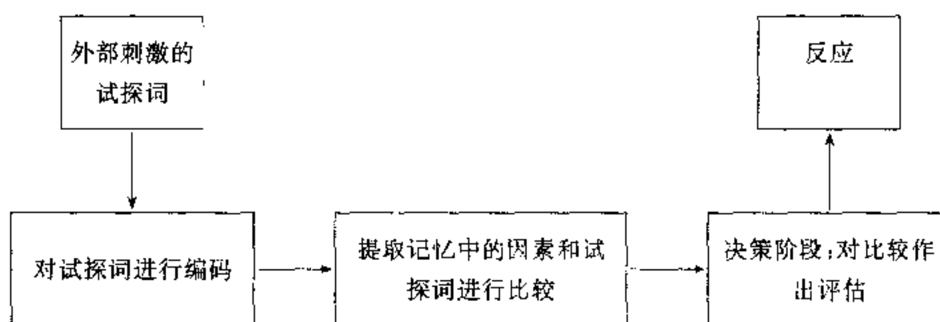


图 9.12 递加因素法示意图

如果每一个阶段都是独立执行的，那么全部反应时（RT）= 编码时间＋比较时间＋反应时间。Sternberg 认为对反应时发生影响的有四种不同的因素：一是试探词的性质（是完整还是衰减）；二是记忆集合的大小；三是反应的类型（“是”还是“非”）；四是每一反应类型的频率。这些因素和不同的处理阶段有关。他把这些因素交叉搭配，发现它们对反应时的影响都是独立的，可递加的。例如他在显示试探词时增加视觉的干扰，便会使全部反应时增加，但这仅影响编码的那个处理阶段，而这个阶段是独立于记忆集合的大小的。这就是说，不管记忆项目的多少，由于视觉噪音的影响，编码的过程都会缓慢下来。

递加因素法为分解一个系列过程的各个成分提供一种精致的、强有力的方法，引起心理学家的兴趣。但是它越强有力，其通用性就越低，因为很多心理过程并不适合这种简单的、线形的、递加的模型。

9.4.2.4. 双任务法

双任务法（dual tasks）的基本出发点是，人的心理（或任何认知系统）的处理资源是有限的。我们完成一件任务所要求的处理资源（注意）取决于这件任务的复杂程度、它需要作出决策的余地的多寡、练习的多少。双任务法用来考察某些任务需要多少处理容量，人们是怎样按照回报来分配资源的。在完成双任务时，次要的任务需要分配较少的资源，因此我们可以用它来间接测量另一任务所需的处理资源。例如我们让被试阅读一篇材料（主要任务），同时要他们注意监听一些随机出现的柔和音（次要任务），一听到音后就要按键，并由仪器将所需的时间记录下来。被试在完成次要任务所需要的时间的多少是他花了多少资源在阅读上面的一种间接的测量。如读的材料很难，有很多不熟悉的概念，人们就需要更多的时间来对付它，而他们监听柔和音的任务就要受影响。

9.4.2.5. 信号监察法

在上述双任务法中，被试必须发现一个微弱的信号，但是在研究知觉和记忆的试验中，较多采用的是信号监察（signal-detection ）的方法。这种方法通常是在显示刺激时，还同时显示或不显示一个信号；显示后，要求被试说出他是否听到一个信号。信号监察法源于统计决策论，因为有信号和无信号都可以产生一些内在的知觉样本，这些样本都是按照正态分布的，而有信号的分布具有较高的平均值。但是这种方法已被广泛地应用在判断试验里，例如 Warren 等通过这种方法发现有一种所谓“语音复原效应”（phonetic restoration effect），被试听到的一句话，The state governors met with their respective legi* latures convening in the capital

city. [州长们在州府和各自的立法机关开会。] 在 * 处，录音被切去 0.15 秒，插入了咳嗽声。所以 legislatures 这个词中间的 s 实际上是听不到的。但是事后询问被试上面那句话中有没有少了一个音素时，二十人中有十九个说没有少掉什么音素。剩下的一个说少了，但在什么地方少的也没有说对。再问他们咳嗽在句子什么地方发生，没有谁能准确说出来，有一半人认为是在 legislatures 以外发生的。甚至当 legislatures 中更多的音素被嗡嗡声所代替，如 l* * * * lature，被试仍听不出来这个词是被打断的。这说明人们在听辩时，并非被动的听，而是主动地对输入的声音进行合成，然后到心理词汇里找寻匹配。

9.4.2.6. 计算机模拟法

作为认知科学的一种较独特的方法是计算机模拟。这种方法的基础是信息处理分析。它根据某种理论或假设，把人类的认知过程变为计算机程序，然后又在被试身上在实验。如果计算机的行为和人类完成相同作业的行为是一致，那么计算机所依据的理论就可以用来解释人类的行为。这种方法的出发点是人类的心理和认知行为往往是一个看不见的“黑箱”或若明若暗的“灰箱”，而计算机和人一样都是处理符号系统（表征）的机制，我们就有可能让计算机这个机制去根据一种理论去设计处理认知的流程，假定这个流程所得到的结果和在真人身上所得的结果是一致的。那么计算机所模拟的流程就有可能是人处理表征的实际流程。Kintsch 和 van Dijk 想了解人们是怎样使用容量有限的短时记忆来将语言输入组合在一起的。如果输入的句子超过短时记忆的容量，人们就只能把前面听过的句子的重要的信息保留下来，并采用它来处理后面出现的句子。保留的应该是前面句子中最为中心的、最近出现的那一部分，如下面的句子：

(1) The Swazi tribe was at war with a neighboring tribe because of a dispute over some cattle. (2) Among the warriors were two unmarried men named Kakra and his younger brother Gum.

[（1）斯瓦兹族因为牛群争端和邻族发生战争。(2) 在战士中有两个未婚男子，叫做克克拉和他的弟弟冈姆。]

第一句子的中心部分是“斯瓦兹族和邻族发生战争”，这一部分和第二句的“战士”的概念联系起来。

又如在下面的句子里，Kakra 是连接两个句子的概念：

(2) Among the warriors were two unmarried men named Kakra and his younger brother Gum. (3) Kakra was killed in a battle.

[（2）在战士中有两个未婚男子，叫做克克拉和他的弟弟冈姆。(3) 克克拉在战争中被打死。]

那么前面出现的概念是怎样和后面出现的概念结合在一起的呢？他们认为有两种机制：第一种是共同的概念标签，如 Kakra 是（2）和（3）句的共同的概念标签。第二种是推断。它是在第一种机制不能起作用时才起作用的。例如（1）和（2）句没有共同的概念标签，就只能推断，第二句的“战士”是斯瓦兹族的成员。

Kintsch 和 van Dijk 把他们的假设编成计算机程序，并把（1）、(2)、(3) 句输入计算机处理。每一个句子都在一个“周期”里处理。在第一个周期里，计算机程序保存了“斯瓦兹族和邻族发生战争”，把“因为牛群争端”废弃。在第二个周期里，计算机程序试图在（1）和

(2) 之间找寻一个共同概念的标签，标签未找到，只好推断“斯瓦兹族”和“战士”是有联系的，并把这种关系保存下来。程序还选择了“两个男子”，“叫做”，“克克拉和他的弟弟冈姆”。在第三个周期，计算机程序又去找一个能把这些概念和第三句的新概念联系起来的标签。这时程序找到了一个，第二句就能和第三句联系起来：“克克拉和他的弟弟冈姆”和“克克拉在战争中被打死”。那么程序的输出和人类的行为是否一致呢？Kintsch 和 van Dijk 的推论是，一个概念在短时记忆中操作的次数越多，就越记得牢，所以在第二和第三周期中出现的概念会比第一个周期中出现的要记得清楚些。计算机用一个概念出现的周期数来模拟一个概念的操作过程数，然后让 100 个大学生做试验，了解他们对这些概念的回述数。结果说明，在计算机中出现的周期数多的概念在被试的回述中的数目也多。反之也少，例如“和邻族”、“发生战争”、“两个男子”的周期数均为 2，而回述数分别为 78，80，82。而“因为”、“牛群争端”、“在战士中”、“未婚男子”的周期均为 1，而回述数分别为 46，39，42，47。

计算机模拟的应用范围甚广，Just & Carpenter (1987) 设计的阅读和理解过程的模拟程序 READER，是一个产生式模型。它把单词编码、词汇提取、进行语义和句法分析、利用图式和建立语篇所描写的世界的指称表征等等过程都结合在一起。READER 一边进行一边建立语篇的表征，因此能够回答与语篇有关的问题并产生一篇小结。这个程序特别注意模拟人们阅读中的时间进程，如碰到难词需要更多的时间处理。READER 通过它所需要的产生式周期来反映了其处理的分量。阅读中关于眼睛固视的研究说明眼睛对一个词的固视反映了人们辨认和理解该词所需的时间。我们只要把 READER 的处理周期和被试所需的时间加以比较，就可以了解模型在多大程度上能说明人类的阅读过程。他们让被试看一篇关于飞轮的 140 个词的文章，得到的结论是 READER 能解释人类阅读 67% 的方差。

9.4.3. 实验方法类型

认知科学的实验方法源于认知心理学，总的来说有两大类：一种叫同时测量法(simultaneous measurement)，即被试在接受刺激时便同时开始测量，例如让被试看一个句子的同时测量其眼睛瞳孔的直径。另一类为连续测量法 (successive measurement)，即在被试接受了刺激以后，才进行测量，例如让被试听完一段话，要求他们把话的内容说出来。有的实验方法只能用一种测量法，也有不少实验方法可按需要使用这种或那种测量法。下面介绍一些主要的实验方法类型：

9.4.3.1. 潜伏性数据

潜伏性数据 (latency data) 即反应时 (reaction time)，认知心理学家历来都对反应速度感兴趣，因为速度的快慢可以反映信息处理环节的多少，知识结构的复杂程度的高低，语义网络中节点的远近等等。过去这些数据只适宜于研究肌动技能，而不是认知能力，因为认知的操作速度甚快，在 1 到 250 毫秒之间，当时的仪器还不够精细。现在使用计算机，准确程度可以在 15 毫秒以内，这就使这种手段大为普及。

潜伏性数据通常用于“次要任务法”。在这种实验方法里，我们让被试完成一件主要任务，同时还让被试完成一件次要任务。例如 Anderson (1985) 认为以知觉为基础的知识表征是线性排列的，可通过观察记忆型式的方式来了解。他让被试学习一些有 4 个辅音组成的字符串，

如 KRTB，这是主要任务。此外还要求他们学习一个与这个字符串相连的数字，如 7，这是次要任务。经过学习阶段以后，Anderson 向被试显示这个字符串，要求他们回述与这个字符串相连的数字。Anderson 感兴趣的回述这个数字的速度，因为这可用来解释他们辨认辅音字符串的时间。在实验中，Anderson 有意操纵显示这四个辅音的次序，例如被试学习的是 KRTB，他看到的有六种不同的情况：

(a) 完全相同	K R T B	1.55 秒
(b) 头两个字母相同	K R B T	1.55 秒
(c) 头一个字母相同	K T B R	1.59 秒
(d) 后两个字母相同	R K T B	1.59 秒
(e) 后一个字母相同	T K R B	1.64 秒
(f) 完全不同	T K B R	1.74 秒

(a) 和 (b) 的时间最短，而且所需的时间一样。其次是 (c) 和 (d)，后一个字母相同的 (e) 和完全不同的 (f) 时间时间最长。这个结果说明记忆有两种重要的效应：一种是向前靠，这可以说是前摄 (proactive) 作用；另一种是向后靠，可以说倒摄 (reproactive) 作用。它不如前摄那么明显。

9.4.3.2. 眼睛固视

眼睛固视 (eye fixation) 的实验方法也叫作跟踪眼睛 (eye tracking) 方法。这种方法收集被试在一段时期里固视某一刺激的数据。这种方法对了解被试视觉的心理过程，特别是阅读过程很有帮助。过去这些数据比较庞杂，难以整理。现在计算机可以帮助心理学家去作大量的数据整理和归纳，所以他们也乐意采用此法。跟踪眼睛固视的原理是，接触眼睛的一小部分可探测的光线从眼睛的各个平面如角膜和晶体的前部反射回来。在测量眼睛固视时，可将一条光线（通常是不会干扰的红外线）投入被试眼睛，并将各个平面的反射记录下来。被试的头部的移动会在数据中产生“噪音”，但是眼睛从各个平面会得到不同的反射型式，可使头部移动和眼睛固视区别开来。一个典型的眼睛固视实验室是这样的：被试坐在椅上，把眼睛投向与计算机相连的屏幕。屏幕的一边是一个记录仪，另一边是一块镜子。把固视的型式反射到光电记录仪上面。

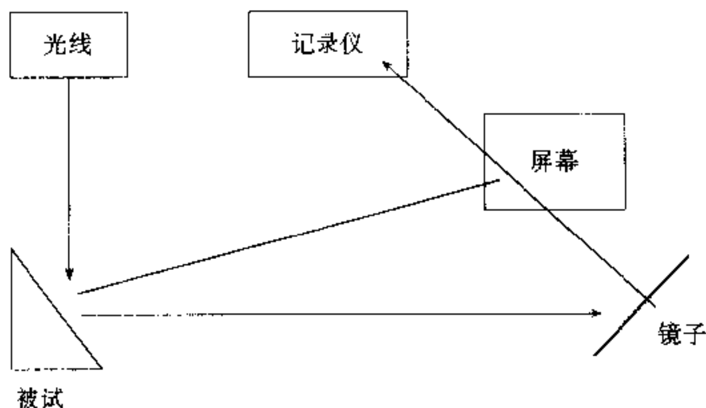


图 9.13 眼睛固视实验室示意图

使用这个方法揭示了阅读中的眼睛移动过程，人们阅读时眼睛并非不断地从一行往一行那样平滑的移动，而是从一个固视跳到另一个固视的扫视（称为 saccade），扫视的平均长度约为 7 到 9 个字母，约为 25 到 50 毫秒。每一固视的时间约为 200 到 250 毫秒。从每一行跳到另一行约需 2 到 5 秒。实际的阅读只覆盖了视网膜的中央凹的中心的一到两个词，在中心以外的区别词的能力就会降低。眼睛固视一个或几个词，使它们结合到语篇所展开的内容里，然后跳到另一个固视点。所以固视时间可用来测量某些词在语篇理解中的难度。阅读能力低的读者自然需要较长的固视，或每个句子的固视需要多些。眼睛移动的记录还表明，人们看实义词的时间长于虚词，看不熟悉的词、不常见的词、意料之外的词要长一点时间，看短语后面的词也要长一点时间。这些记录也显示，读者的眼睛会回溯语篇中较难理解的一部分。如果是歧义的句子，在理解有困难时，他们会回溯到歧义的分叉点。例如读到 The boat floated down the river sank［顺流而下的小船沉没了］这样一个句子，读者会以为 floated 是句子中的主要动词，但到最后发现另一个主要动词 sank，他就会回溯到句子前头，重读句子，把 floated 理解为修饰 boat 的形动词。读者在读到代词或未指明对象时，也会到前面去搜索先行词。

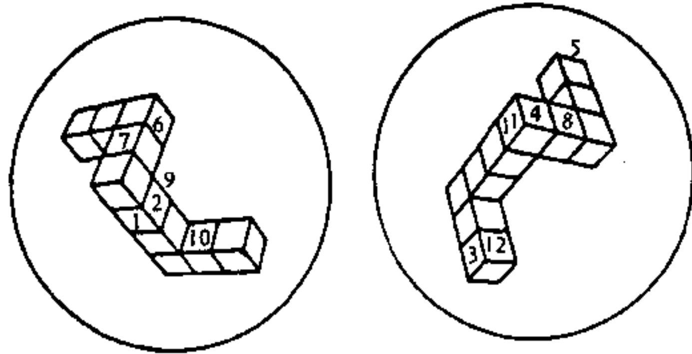


图 9.14　心理旋转试验

眼睛固视的方法还可以用于许多心理感知的问题，例如 Just & Carpenter 曾用这个办法来研究空间表象：他让被试看一些三维物体的两维表征，要求他们作心理旋转，看两个物体是否匹配。被试作判断时眼睛固视的情况记录下来，实验说明被试眼睛先找出主要的轴心（下图中的固视编号 1 和 2），作心理旋转，然后检查他们的判断是否正确，先看物体的尾部（编号 3-5），然后才是别的部分（编号 6-7）。实验结果证明人们在处理空间表象时，在内心里模拟实际情况进行旋转。

9.4.3.3. 口头报告

口头报告（verbal report）是早期心理学研究中内省心理学家爱采用的方法，但是用这种方法所收集的数据往往有很大的任意性的成分，在行为主义心理学家的攻击下，这种方法被摒弃了半个多世纪。随着认知科学的兴起，这种方法得到复苏，越来越多人试验这种方法，因为它可以收集大量数据去检验那些需要解释每一分钟的行为的模型。

口头报告有两大类：

1．共时报告—有声思维

过去内省心理学家把被试的表面的口头陈述看成是他本人的思维过程，而现在的口头报

告则用来说明实验者所生成的理论(往往以计算机模拟的形式出现)的数据源。Ericsson & Simon (1980)认为,我们必须区分那些导致可信而有效的口头报告和导致不可信而无效的口头报告的条件。作口头报告和作其他认知心理学试验一样,对容量有限的短时记忆、对长期记忆和推论过程都有一定的要求,这些要求取决于口头报告本身的性质。例如我们要求被试边做事边报告他们在做什么,就会增加短时记忆的负担,因为他们要同时注意两件事情。因为负担沉重,被试就只好想法来对付实验的要求。这种方法用处不大。但是有另一种口头报告,可称为"有声思维"(thinking aloud)。在有声思维的实验中,我们不要求被试报告他们是怎样想的,只要求他们报告在想些什么。换句话说,他们只需报告他们在做一件事情时的短时记忆中一闪念的思维活动。

有声思维的理论认为,生成口头报告的过程是生成所有可观察行为的一部分,可作信息处理分析。信息有序地进入短时记忆,要注意刺激需要时间,这就占去了短时记忆的容量。这种理论的主要特征是被试只需报告其短时记忆的内容。

有三种类型的口头报告:

(1) 直接表达,把已经在短时记忆里用词语编好码的信息简单地表达出来。被试只是把记忆里的东西说出来,说话并不会影响所要完成的任务。被试在解决问题是对自己的自言自语属直接表达;在实验里无非是要求被试"自然地有声说出来",避免内省和报告自己的思想,避免追求完整,也属直接表达。

(2) 对短时记忆的内容进行编码。用口头报告描述短时记忆中尚未用词语编码的内容。要求被试在完成与表象有关的作业的同时进行有声思维,他们可能在作出口头报告前就要把记忆中的一些内容进行编码。这种编码的过程会和完成作业的过程争时间。因此有声思维会使完成作业的时间延长,但不一定影响处理的程序。如果处理作业的负担过重,口头报告就会停下来。

(3) 解释。要求被试解释他们的思维过程,他们就要进行综合加工,这往往会导致作业的中断。被试的解释不一定是作业过程的准确的描述,这要看是什么作业,什么样的被试和所作的解释的性质而定。例如实验者从旁提示,"你现在在想些什么?"就能引起被试解释其思维过程。

这种有声思维的口头报告往往用录音的办法来进行记录,有时还需要从旁观察,工作量较大,一般只能做个案研究。被试往往对有声思维的要求往往不很明确,需要先做训练,否则报告出来的不是他在想什么,而是他是怎样想。例如下面是一个类比的题目,要求被试报告他在解题时想些什么:

"华盛顿比林肯相当于1比________
a. 5　　b. 12　　c. 13　　d. 22"

典型的有声思维大致是:

"华盛顿比林肯相当于1比________
……华盛顿……第一任,……5,不,……22,不,……13。"

这说明被试首先认出他们都是美国总统，然后想到华盛顿是第一任总统，再去答案中找相当于林肯为第几任总统的数字，他排除了 5 和 22，最后决定为 13。其实林肯为第 16 任总统，而正确的答案应为 5。被试一开始就没能抓到正确的类比。在美国的纸币里一元面额的纸币上印有华盛顿的头像，而五元的印有林肯的头像。

如果被试报告的是，"首先，我在比较华盛顿和林肯，看他们有些什么相似的地方……"，他就不是在报告他在想什么，而是报告他是怎样想的。

2. 回顾式报告。要求被试在解题后报告他们解题时的思维过程，这种报告可以弥补共时报告的空白，可以提供被试的全面的看法。但是对报告的内容的分析必须小心，因为他们很可能重构一些在解题时并没有出现的事情。所以要注意两个问题：(a) 解题后所做的报告越延缓，就越有可能是重构和歪曲原来的思维过程；(b) 回顾式报告提出一些可疑的独立的证据，需要用别的数据来支持，如反应时和错误分析。

这两种类型的口头报告并非互相排斥的，可以要求被试做两种报告。口头报告对了解学生的认知过程和学习过程很有帮助，例如王初明、亓鲁霞（1992）使用英语听写作业来了解英语水平不同的被试的学习策略，要求他们既做共时报告，也做回顾式报告，以互相引证。他们的研究发现，水平高的学生较多采用联想的策略，而水平低的学生，认知负担过重，只能把注意力集中在语音的形式上面，使他无法发挥从上而下的信息处理策略，包括联想的策略。

9.4.3.4. 双耳实验

双耳实验（dichotic experiments）使用特制的耳机进行，这种耳机的两个频道可传送不同的信息。这种实验可用来了解听觉处理信息的模型，也可用来了解句子听辨的许多问题。早在 50 年代实验心理学家 Broadbent（1954）使用了这个办法来验证他的过滤论（The filter theory），后来 Treisman（1960）又使用它来验证她的衰减器模型（The attenuator model）。在 Broadbent 的实验里，被试两耳同时分别听两组数字：

	左耳	右耳
第一组	7	3
第二组	4	2
第三组	1	5

每一组的数字的间隙是半秒，被试听完这三组数字后必须把听到的数字说出来，次序不论。结果是被试大都先说一个耳朵听到的数字，然后才是另一个耳朵听到的，准确率为 65%。如果要求被试把一组组的数字说出来，准确率就降低至 15%。实验说明每一个耳朵是一个知觉频道，当某一个频道受到注意时，要转移到另一频道需要时间。因为每一组数字的间隙只有半秒，被试来不及转换频道，只好把每一组的刺激放在一起处理。但是后来有的实验的结果表明，未受注意的另一个频道也能听到一些信息，这是过滤论难以解释的，于是 Treisman 让被试的两耳听不同的句子：

右耳："I sam the girl song was wishing..."
左耳："Me that bird jumping in the street..."

这两个句子都有一部分是没有意义的，但把右耳的上部分和左耳的下部分联系在一起却成为有意义的句子：I saw the girl jumping in the street. Treisman 发现被试在跟读这些句子时会自动转换频道。这说明被试在注意一个频道时，另一个频道的信息并没有消失，它只不过是衰减而已。

9.4.3.5. 辨认与回述

这两种手段都是用来了解记忆的。

在辨认（recognitition）的实验里，要求被试辨别一系列的检测项目（如文字、声音、图形等），判断每一个项目是否在实验中的习得阶段出现过。换句话说，他们必须区别哪些是原来出现过的探测（probe）项目（目标项目），哪些是没有出现过的项目（干扰项目）。在辨认里，我们仅要求被试回答是/否；在再现里则要求把项目说出来或写出来。有时我们在辨认里还加上一个信心量表，让被试选择他对自己的回答的把握程度有多大。有的辨认使用多项选择的格式，让被试在几个被择中选出答案。选择题中的干扰项和答案的相似程度越大，干扰性就越强，例如要从 48.0，3.1416，6.3842 几个项目中辨认 p 值比从 3.4116，4.3146，3.1416 中辨认要容易得多。Sachs 给被试看一个故事，其中一关键句是主动句或被动句。故事后是一个辨认的实验，要求被试指出一个句子是否故事中的原句还是经转换过的句子。结果表明，如果辨认的句子是最后的句子，被试都能认出来，如果句子是在材料的前面，辨认率就只有 60%。

在回述（recall）的实验里，被试必须根据一些提示来产生一些以前学习过的项目。虽然它也和记忆有关，但要求与辨认相反：在辨认里，我们向被试提供一个项目，被试必须回忆它和习得语境的联系；而在回述里，我们向被试提供习得语境的联系，被试必须回忆与此联系的项目。有人通过要被试回述一段材料的内容来研究语境对理解的影响，发现语境的作用是在编码过程中，而不是在回述时发生的。有人有故意改变语境出现的时间，发现语境出现在文章之前的作用要大于出现在文章后的作用。又有人改变文内的单词概念的数目，发现其对回述有影响。桂诗春和李崴（1992）使用了句子回述实验来研究短时记忆中意义的作用。他们在学英语的中国学生两种句子的对比实验。被试首先看到是一个汉语的句子，如：

“罗队长下令在援军到来之前要死守防线，不得后退一步。”

听完句子后，被试还要在屏幕上看一个有 5 个词的词表，这几个词是逐词显示的，一个代替一个。

汽车　菊花　茶叶　阵地　帝国

然后再给一个词如“光明”，要求被试判断这个词是否属于前面的词表里的词。（这一部分为辨认试验）值得注意的是，词表中有一个词“阵地”是有诱惑作用的。最后被试还要把句子回述。实验的设计思想是，如果按照传统的说法，短时记忆保存的是句子原来的表层形式，那么要求被试作原词回述，他们只能回述原来的句子，词表中的诱惑词不会起诱惑作用（即意

义不会介入)。如果被试在回述时，把句子写成“罗队长下令在援军到来之前要死守阵地，不得后退一步。”就说明意义在短时记忆中还是起作用的。实验结果表明，不但在英语句子+英语词表的状态下，就是在英语句子+汉语词表的状态下，诱惑词的介入率都比较高。意义在短时记忆中的作用是明显的。

9.4.3.6. 判断

判断实验也是经常用来研究记忆的方法，例如可让被试判断两个项目中哪一个在习得阶段中出现得较多（频率）或出现的较近（近现率)，哪一个是大写或小写，哪一个是法语或英语等等，这种记忆判断实验能够展示一些辨认和回述难以取得的事实。判断实验也可用来研究人们的世界知识的结构。例如 Rosch 让被试用一个 1 到 7 的量表，来对一个范畴里的各个成员的典型性进行评分，1 表示非常典型，7 表示非常不典型。被试的评分是相当一致的。在鸟的范畴里，“知更鸟”为 1.1 分，而“鸡”为 3.8 分。在体育运动范畴里，“足球”为 1.2 分，而“举重”为 4.7 分。在罪行的范畴里，“谋杀”为 1.0 分，而“流浪”为 5.3 分。在蔬菜的范畴里，“胡萝卜”是 1.1 分，而“欧芹”为 3.8 分。这说明一个范畴里的成员的区别是程度上的问题，没有绝对的界限。

判断实验也可以用来研究说话人的语言知识和他们对御用的期望。例如我们可以让被试判断哪一类词可以放进一个句子框架里，使句子成为一个可接受的句子。我们也可以让被试判断两个句子的意义的相似程度，例如我们想了解句子转换后是否保存了原来的意义。就可以问被试 Was that John? 和 Wasn't that John? 的意义是否一样。同样的，我们也可以就句子的语法性和适宜性要被试来判断。有的证实实验(verification experiment)也是一种判断实验，为了证实已知信息和新信息，Hornby 向被试读一些句子，然后要求他们判断句子是描写哪些图画的。实验说明被试是能够区分已知信息和新信息的。(见桂诗春，1991)

9.4.3.7. 转移

转移实验可用来考察以前获得的知识和能力是怎样转移到新知识和新能力的学习上面的。例如有人让被试学习一种文字编辑程序，待他们能熟练掌握后，又让他们学另一种文字编辑程序。然后从中观察，发现被试在学习第二种程序时，在时间上大大缩小，可转移的不但包括一些指令，而且是更重要的是许多有关的基本概念，如文字编辑器能干些什么，文字编辑的步骤，文件的生成、提取和存储，等等。学习第一个程序可以帮助掌握第二个程序（前摄作用)，而第二个程序又反过来改进和巩固第一个程序的一些共同的东西（后摄作用)。

但是不是所有的转移都是对双方有利的。如果这两个文字编辑器的键盘系统和指令完全不一样，那就会出现负转移。如果经过一段时期，被试再上机操作文字编辑器，就会出现遗忘。后面学习的资料干扰了前面学习的资料，叫作前摄干扰。前面的干扰了后面的，叫作后摄干扰。

从认知心理学的角度看，转移实验可用于研究知识结构、图式、语义网络、概括等认知活动；从语言学习的角度看，转移实验可用于研究语言知识和语言能力的关系。特别是在第二语言习得的研究中，母语和另一种语言的关系究竟怎样？这可说是一个永恒的研究课题。

9.4.3.8. 概念学习实验

一般来说，关于辨认、判断和转移的实验大多数和硬记有关，即了解被试从记忆中提取和原来编码相同的形式的能力。这些试验固然很重要，但是并非研究学习和记忆的唯一的实验方法。例如归纳和概括的思维过程也和学习和记忆有关。这就是概念学习实验。

一个概念就是对一些型式或环境中的共同特征的内部归纳。一个概念也可以说是把成员和非成员区分开来的“决策规则”。作为一个特定概念的范例的刺激集合就是范畴(category)。我们关于狗的概念就是我们认为狗应该是怎样的一个模型；而“狗”的范畴就是用这个模型来表示的一个物体集合。哲学家把前者称为内涵，后者称为外延。

在概念学习实验里，我们在实验室里观察被试是怎样学习一些简单的人为的概念，从而了解被试在不同环境中进行归纳式学习的一般原则。第一步是训练阶段，教被试怎样把一些刺激区别为一个范畴里的成员和非成员。例如向他们展示一些刺激，让他们决定其是否一个范畴的成员，实验者从旁给予反馈，告诉他们其决定是否正确。这个过程反复进行，一直到被试掌握为止。第二步是转移阶段，给被试一些新的刺激来区分。被试怎样区分刺激有助于我们了解他们在训练阶段所学到的概念模型（规则)。这两个阶段不一定明确地分开，可以在训练阶段的若干地方安插一些被试不熟悉的新刺激，以检验被试掌握概念的情况。例如让被试对一个范畴里的新事例和旧事例作出反应，可以了解被试是在硬记还是在作归纳。如果被试能够对在训练阶段熟悉了的事例作出正确反应，而不能对新事例作出反应，就可以说他使用了硬记的方法来学习。Santa（1977）用这种方法以观察人们以知觉为基础的知识表征，发现像几何图形在记忆中是按其空间位置保存的，而言语则是按线形次序保存的。

还有一种和概念学习实验相近的实验方法是以知识为基础的学习实验，主要考察人们对于现实世界的物体、事件、社会群体等等范畴的知识结构的学习过程。它所考察的是人们已经具有的知识，不像前一种方法那么容易控制，但它所了解的是人们在现实生活中的一些更为复杂的概念。这种研究通常需要增加两个或更多的步骤。首先，我们要系统地收集被试关于所考察的概念的直觉的知识，例如让被试列出一些活动（如上饭馆吃饭、去自动洗衣店洗衣服）的典型事件，或是让他们在一个量表（如概率程度、与中心事件的相关程度，等等）上给这些典型事件评分。然后再用这些评分来建立不同物体的范畴，例如相关程度低和相关程度高的事件，最后是观察被试在关于这些物体的记忆实验中的结果和评分结果有些什么相关。一种典型的实验方法是让被试看几段描写不同的定型事件，然后再进行回述或辨认。被试在记忆实验中所犯的“错误”特别有意思。例如回述的是上饭馆吃饭，被试可能记不清楚原来的描写，就会错误地回述一些他们认为极有可能出现的典型事件。这些结果可以提供线索，帮助我们了解被试是怎样认识语篇、怎样把新知识和旧知识组合、怎样使用知识来重建记忆不清楚的事件。在找出了典型事件的中心点或典型性以后，我们还可以进一步观察其轨迹，例如可发现它出现在被试的编码过程，因为他们较多地注意某一类事件；或是发现它出现在被试的检索过程，因为他们利用了一般的知识来填补记忆中的空缺。

9.4.3.9. 在线测量

在上面谈到的一些实验方法已包含了在线测量的成分，如眼睛固视，一边在看，一边在测量眼睛移动。下面再集中介绍几种常用的方法：

1. 跟读 (shadowing)。让被试带着耳机，跟读听到的声音。在双耳实验里，两个耳朵听到是不同的内容。在别的实验里，听到的是有白噪音干扰的信息，听到的单位从一个音素到音节，一直到单词和句子。这种方法可用来研究言语听辨的各种现象，如信噪比的高低和听辨的关系，一般的看法是，信噪比越低，听话人所能检测的信号区别就越小，它接受信息的可能性就越微薄。

2. 阅读时间。这种方法被广泛地应用在了解阅读理解过程，一般用来测量被试阅读一个语言单位（单词、短语、句子、语段）所需的时间。随着计算机的普及，让被试在屏幕上阅读并记录时间，已经是轻而易举的事情。被试在计算机上阅读，可以通过按键盘来控制显示材料，逐词显示，这种方法称为快速系列视觉显示（RSVP，rapid serial visual presentation）。这些词可以在屏幕上同一地方出现，也可以顺序在一行中出现，好象是在阅读一句话一样。在RSVP 中测量的是阅读每一个词所需要的时间，也就是被试第一次按键到第二次按键之间的时间。时间的快表明被试在处理该语言单位时比较顺利，慢表明处理上遇到困难。例如音节多的词、不熟悉的词、在语义上预料不到的词、出现在预料不到的句子结构里的词、一个主要成分结尾的词，都需要较多的时间。

也可以用同样的方法来测量阅读每个短语和句子的时间，以了解一些语言结构的变量，如：句子的音节长度、被试对句子中的单词的熟悉程度、句子的复杂程度等等。阅读时间也可以用来了解期望在语境中出现的句子的作用。RSVP 的方法在在线实验中被普遍使用，因为它对实验变量很敏感，而且在计算机上很易实现。读者如对这方面感兴趣可参阅 Just & Carpenter（1980，1987），李绍山（1992）。

3. 试探反应时（probe reaction time）。在阅读中，读者在处理一个句子的较难部分，例如碰到句法上复杂或语义上歧义的句子，或代词的前指不明确时，需要更多的资源。我们往往想测量被试处理这些问题时究竟用了多少注意力资源，就给被试双重任务。如主要任务是用 RSVP 方法进行阅读，而次要任务是让被试注意听一些偶尔出现的柔和音（或看一些偶而出现的暗光），一旦它们出现就要按键。这种试探性的刺激并不常出现，而且没有什么规律性。如果主要任务吸引了被试较多的注意力，他的试探反应时就会较长。这是因为我们短时记忆的容量有限，主要任务已用去它的大部分资源。

4. 音素监听（phoneme monitoring）。这种方法让被试在听一段语流，同时注意监听一个音素，如 (ba)，要求他们一听到这个音素时，就按键。例如被试听的是一个句子 The emperor went to royal baths，当他们听到 baths 中的 ba 时，就必须按键。例如 Foss 让被试监听歧义和非歧义词（如下面句子中的 straw 和 hay）后的音素/b/（beside 中的第一个字母）：

The $\left\{\begin{matrix}\text{farmer}\\\text{merchant}\end{matrix}\right\}$ put his $\left\{\begin{matrix}\text{straw}\\\text{hay}\end{matrix}\right\}$ beside the machine.

straw 有两层意思，一是“稻草”，一是“吸管”，而 hay 只有“稻草”的意思。但在第一句里，farmer 可提供语境来解决 straw 的歧义。尽管如此，被试还是受了歧义的影响，因此需要更长一点时间来对/b/作出反应。使用这种方法的出发点和试探反应时的出发点是一样的：理解的首要任务占用了被试短时记忆的处理时间越多，监听就会越慢。音素监听是一个较为复杂的过程，一些无关的变量往往会起到干扰的作用，例如监听的音素在不同句子中的频率和分布

不同对监听也会发生影响，一些音素的区别度和响度不同，被试对音素所在的单词的熟悉程度不同，也会起作用。

5. 插入问题（interrupting questions）。要了解阅读过程的另一个简单的方法是在被试接受语篇时，向他提出问题，打断其过程。所提的问题可以是和前面的语篇有关或无关的。例如要被试说出他刚听到的句子中的代词的指称对象，并测量他作出回答所需的时间和准确程度。另一个例子是在听完或读完一个句子后，立即打断被试，向他提出一个关于谁/什么/哪里/什么时候的问题。这种方法可以用来检验各种假设，例如一些比较复杂的句子结构中的因素较难提取，所以听了主动句后提供一个答案要比听了被动句后容易些。另一种插入的问题可用来考察前面的句子或语篇怎样激活长期记忆中的语义信息。例如在词汇决定的试验里，让被试决定一个提示字符串是否一个词。被试看到的有一半是正常的词，而另一半是像正常的词的非词（如 order 和 ordar）。决定一个词是否真正的词所需的时间取决于这个词是否被试所期望的或是否为前面的语境所激活的。例如语境是到饭店吃饭，order［点菜］就要比 older［年纪大一点］所需的决定时间要快 50 毫秒。在心理语言学的许多研究中，词汇决定试验是一种收集信息的重要工具。试验表明，在听了或读了一些出现在一定语境中的多义词后，被激活的是几个意思，而不仅是与语境相适应的那一个意思。例如听了 John deposited his money in the bank 后，被激活的不但有与语境相关的词（office［办公室］，saving［存款］，teller［出纳］），而且还有无关的词（river［小河］，water［水］，ground［土地］）。这种效应是和时间有关的，紧接在多义词后的无关的激活是可测量的，不过经过一秒多钟的间隙后，有关的激活就占了主导，无关的激活就衰减了。（Swinney，1979）

9.4.4. 话语分析的实验方法

在上一章里，我们介绍了话语分析的定性和定量方法。比句子高一层次的话语分析不但是一个文化语言学、语用学、语料库语言学、心理语言学的交叉领域，而且也引起认知科学家的极大兴趣，因为这是人工智能研究不能回避的一个课题。由于大家从不同的角度来看话语分析，因而所采取的研究方法就不尽相同。认知科学家采用了实验的方法，特别是计算机模拟的方法，因为他们的目标是实现人工智能和建立自然语言处理系统。他们研究的中心是了解那些使话语连贯在一起或不一起的机制，了解语境是怎样影响人们使用和了解各种语言表达式的。下面我们主要是根据 Grosz 等（1989）从四个方面来考察话语分析的问题以及所使用的研究方法。应该指出的是，我们所介绍的看上去象一些模型，属理论上的探究，但是它们都是在计算机上实现的模型，不过具体的技术处理和算法难以在这里讨论。这些模型体现人工智能科学家使用计算机来处理语篇的一些努力和尝试。

1. 有关语篇结构的问题
2. 语篇结构的语言指示器
3. 有关短语层面（如代词、特定的描写、省略式）的问题
4. 有关语篇计划的问题

9.4.4.1. 早期的研究

早在 70 年代就有人从建立以计算机为基础的完整的自然语言处理系统的角度来研究语

篇的计算方法。这些系统不像现在那样有一个话语结构的模型。但是系统的研制者采用实验的方法来了解代词的使用过程和有限度的意向辨认过程。Woods (1972) 所建立的 LUNAR 系统提供了一个解决代词的指称对象的机制，但是这个系统只能进行孤立的一对对的问答，话语的模型很简单：只把前面出现过的实体保存下来，以了解代词所指的对象。Charniak (1977) 企图了解在儿童故事中识别特定描写和代词的指称对象，其方法是用推理规则和有关的启动程序来对某些领域的信息进行编码。Winograd (1971, 1972) 的 SHRDLU 是一个关于积木的对话程序，它可以理解用户的意图，执行一些移动积木的指令，并对动作保存记录，它可以解释一些代词和特定描写，像“一个”类型的前指式。

70 年代后期第二代自然语言系统集中在研究语篇的特定领域的知识和怎样把代词以外的语篇现象组合在一起的方法。大多数的系统都把语篇的领域局限在范围很小的活动和事件。但是 Lehnert (1977) 试验建立一个能够理解就故事内容提出非线性问题的系统。Grosz (1977) 等开发的任务会话理解系统 (Task Dialogue Understanding System) 把领域知识（表示为一个任务模型)，语篇信息（表示为整体的焦点和解释指称的有关算法）和识别意图（处理有关任务模型的问题和更新任务模型）区分开来。这是第一个考虑到语篇的（语言）结构和对语篇中表达式的解释的相互作用的系统。早期的研究使研究者认识到语篇处理不能看成是一些简单的了解指称、辨认意图和使用领域知识的程序集合，我们需要一种把这些过程区分开来的手段，并对它们和它们之间的相互作用提供一种机制。

9.4.4.2. 语篇结构

一个语篇的语句并非杂乱无章的，就像句子里的单词由句法结构组织起来一样，语句亦有自己的结构。语篇可以分为语篇片段 (discourse segments)，而这些片段相互之间又有不同的关系。在语篇类型分析中，受到人们注意的有：以任务为目标的会话、复杂对象的描写、叙述性文体、正式与非正式的辩论性文体、商洽性文体和解释性文体，等等。对语篇结构的认识有助于了解语篇意义理论和语言处理。语篇意义理论的核心问题是规定语篇的基本单位及其相互之间的关系，我们必须在语篇处理中决定哪一个语句和语篇的哪些部分有联系，所以语篇结构在语篇处理中的作用体现为它能界定语篇的意义单位，限定哪些语篇单位对解释某一语句的作用。

早期的语篇理解研究（如 van Dijk 和 Rumelhart）建议采用一种相当于句子语法的语篇（故事）语法，而以任务为目标的会话则认为会话的结构取决于正在执行的任务的结构。后来的研究则采取下列的方法中的一种：(1) 采用语法的概念；(2) 规定一些修辞的或语篇的关系作为语篇结构的基础；(3) 考察作为语篇结构的来源的特定领域知识或普通常识；(4) 把广义的交际意图（任务结构的一种概括）和它们之间的关系作为语篇结构的基础。具体来说，对语篇结构可从语言结构、修辞和交际意图三个不同角度来进行观察：

1. 从语言结构的角度看语篇结构。这是语篇分析常用的一种手段，它考虑的主要是这样的问题：哪些语句组成一个语篇片段？这些片段之间有些什么关系？例如 Linde (1979) 和 Polanyi (1986) 所提出的模型体现了社会语言学的传统，从表层行为的角度来解释语篇，Polanyi 认为语篇的等级结构“来自说话人用以建立他们的语篇的语言单位中的结构和语义关系。”在她的模型里，语篇树结构由一套语篇语法建立起来。子句（以及和它们相联系的语义）是树结构的节点。换句话说，语篇结构是一个以语言因素为节点的树结构。这些理论的

特点是集中在考察语言结构，而避免在模型中提到交际意图。

2. 从修辞的角度看语篇结构。这些研究的特点是强调话语之间的意义，虽然它们也用语言节点来建立树结构，但是焦点却是语句之间的意义联系（语义关系）。Cohen (1987) 分析了各种辩论体，并使用了提示短语 (cue-phrase) 信息和推断证据相结合的方法，以决定辩论体的结构。有不少研究提出以修辞关系作为解释语句和话语片段之间的关系的基础，每一种关系都必须处理特定领域的信息，以决定这种关系在语句处理中是怎样辨认和产生的，例如 Hobbs (1979) 定义了一套连接语篇片段关系的手段（如平行、实现、对比等等），连接两个语句之间的关系取决在特定领域的事实基础上所作的推断。Lehnert (1981) 所提出的各种关系来自心理状态和事件的模型，她认为可用一套基本的“情节单位”(plot units) 来描写各种可能的状态或事件之间的过渡，这样一来就有可能对特定领域的信息进行编码。这些情节单位就可以标志语篇片段之间的关系。又如在 McKeown (1985) 的关于语篇生成的研究里，提出了一系列的修辞性谓词（包括属性、识别和比较），这些谓词组成各种图式，以定义不同的语篇类型。

3. 从交际意图的角度看语篇结构。Grosz & Sidner (1986) 认为，语篇结构的根源应该是以计划为基础 (plan-based) 的关系或交际意图的关系，而语言结构只起到一种表示内嵌关系的作用。语篇片段之间的内嵌关系部分地取决于片段的某些语言特征（如语调或提示短语），部分地取决于语篇片段所传递的意图。语篇片段层面的意图并非简单的语篇层面意图的一个函数，而是一个由语句，特定领域事实，语篇意图，以及对它们的推断所共同组成的一个复杂的函数。他们认为语篇结构由三方面的相互联系的结构组成：除了语言结构以外，还有意图结构和注意状态。

意图结构包括语篇片段各种目的和它们之间的关系。语篇片段目的体现话语参加者的意图，它导致话语的产生。任何想到的东西都可以成为语篇片段意图。在许多语篇片段意图中有两种关系是常见的：支配关系 (domination) 和满足优先关系 (satisfaction precedence)。这些关系反映了两种事实：一是满足了一个意图导致另一个意图的满足；二是必须先满足一个意图才能满足另一个意图。语篇片段目的部分地取决于对这两种关系的认识，部分地取决于特定领域的知识，部分地取决于语境的其他特征。

注意状态反映了话语参加者在语篇进程中的注意焦点。每一个语篇片段都有一个焦点空间，这些空间构成一个先进后出的堆栈。当片段进入语篇时，堆栈就膨胀；当其意图得到满足时，堆栈就收缩。这个注意状态的焦点空间模型制约了语篇的处理过程：它规定了什么时候使用什么语言表达式，并决定某一个语篇片段目的在什么时候起支配作用，或起满足优先的作用。

9.4.4.3. 语篇结构的语言指示器

如上所述，要理解一个语篇必须首先认识它的结构。但是结构是怎样认识的？意见远非一致，大家都承认，说话人具有一些强有力的语言机制，以帮助听话人理解一段正在进行的语篇。这些机制（如提示短语、语调型式、姿势）可以指出语篇片段从什么地方开始和在什么地方结束，它和别的片段是怎样联系的。

1. 提示短语亦称为提示词、语篇标记、语篇小品词。提示短语指的是一些表达式，它们对语句的意义没有直接影响，但却传递了该语句的语篇结构的信息。像英语中的 now [现

在]，in the first place［首先］，by the way［顺便说］，汉语中的“好吧”，“可是”，“我想起啦”。但是提示短语究竟标志了什么结构信息，却要看我们是站在什么角度来理解语篇的。例如那些把语篇结构主要看成是语言结构的人认为提示短语直接标志了语言结构的等级中不同的片段是怎样联系的。那些认为语篇结构具有一种底层的修辞关系的人认为提示短语标志了两个片段间的一种特定关系，如 That is［这就是］表示“增添”，similarly［同样的］表示“平行”，for example［例如］表示“例证”，but［但是］表示“对照”。那些强调语篇结构中的交际意图和注意的作用的人则认为应该从他们所提出的三种互相联系的结构的模型来对待提示短语。例如像 that reminds me［这使我想起］和 anyway［不管怎样］表示注意状态的改变，前者表示进入一个新的焦点空间，而后者表示跳回到原来已经建立的空间。而 incidentally［顺便提一下］则说明说话人要偏离话题，因此意图结构就要延伸到一个新的意图等级。虽然从不同角度看语篇结构的人都认识到提示短语的作用，但是大家也承认，提示短语并非必不可少的，而它本身也不足以决定语篇结构，有不少语篇结构并没有提示短语，有不少有提示短语的语篇也只能对各种可能的结构提供一些制约。归根结底，语篇结构还是取决于它所包含的语句所传递的信息，取决于这些信息是怎样联系在一起的。而提示短语只不过是更有利于决定这些联系。

2. 语调和姿势。语调是标志语篇结构的有力工具。有些研究表明，在自发产生的语篇中的停顿长度的变化、口语的速度和结构片段的分界是相关的。也有的研究指出，某些语调特征和 Grosz & Sidner 的模型的某些成分有密切的对应关系。音高的范围可以标志语篇片段的分界，重音可以提供注意状态的信息，声调可以标志意图结构。姿势也是一种标志语篇结构的有用的工具；在面对面的会话中，尤为重要：它不但可以标志语篇片段的分界，还可以提供关于注意焦点和意图结构的信息。

9.4.4.4. 短语层面现象

短语层面现象（phrase-level phenomena）指的是理解语句中个别短语或子句时所发生的问题，这些问题大都和语境有关。例如代词、特定描写（definite descriptions）[1] 的理解和生成都明显地受到语境所影响。又如修饰词的依附和语句中的省略，亦属这类现象。从 Grosz 等人的模型来看，注意状态也可用来解释短语层面现象。

1. 代词和特定描写。人工智能和机器翻译的的研究最关心的是怎样决定话语中的代词和特定描写的指称。一个问题是决定有哪些可能的指称对象。另一个问题是怎样从中选择最合适的一个。要解决这个问题有两种方案：一种是把决定指称对象归到更为一般的推断过程，如 Hobbs；另一种是从指称对象怎样和注意状态相互起作用的角度去解决，如 Grosz。第一种方案只考虑指称对象的解释的问题，而第二种方案则还要考虑规定那些足以影响生成合适的指称对象的制约。

在第二种方案中，焦点和聚焦过程在解决指称对象中起了重要作用。聚焦指语篇参加者在语篇进程中注意力焦点的移动，焦点又分为整体的和局部的两种。整体的焦点以焦点空间的堆栈为模型，对特定描写的使用和解释发生影响。局部的焦点以中心和形成中心（centers and centering）为模型，某一语篇片段的中心是注意状态的一个要素，每一个新片段的开始都

① 和定冠词一样，特定描写指的是一些有指称的描写性词语。

会引入新的中心。形成中心对代词的使用和解释发生影响。

整体的焦点堆栈的每一个空间都包含了参加者在相应的语篇片段中，乃至语篇片段目的中所注意的客体。处于焦点中的客体（即在焦点堆栈中的某一空间）是特定描写的指称对象的主要候选。它们也是其他隐含的客体的源泉（例如当焦点是“书”，“皮”就会用来指“书皮”，而不是其他的皮）。虽然注意焦点对了解前指式至关重要，但是那些首先触发前指式所指的客体的短语和语句对了解前指式也是十分重要的，有人称之为“触发话语实体的描述”(discourse entity invoking description，简称为 ID)。

2. 事件指称。事件指称引起的的问题有两种：一种是找出前指式所指向的事件，一般是采用代词，如 John runs every day of the week. That's his main form of exercise.［约翰一周内每天都跑步。这是他的主要运动方式］That 等于他每天都跑步。也可以用一些包含有“做”的成语，如 John runs every day of the week. He does it to stay healthy［他这样做是为了保持身体健康］。第二种是决定话语所描写的事件在时间上的相对次序，影响次序的一般是动词的时态、体或副词修饰语。注意状态的变化也会起作用。例如下面几句话：

(1) John went over to Mary's house.［约翰上了玛丽家］

(2) On the way, he stopped by the flower shop for some roses.［在路上，他在花店停下来买些玫瑰］

(3) He picked out 5 red ones and 3 white ones.［他挑了五朵红色的，三朵白色的］

(4) Unfortunately they failed to cheer her up.［不幸的是，它们不能使她高兴］

第一句引入发生在过去的事件 E1，第二句描写发生 E1 前的事件 E2，第三句的事件 E3 发生在 E1 前和 E2 后，但发生在 E2 的地点。第四句的事件 E4 发生在 E1 后。这在计算机自然语言识别中会带来很多问题，必须有一个处理时间上先后次序的模型。Webber (1987) 提出以焦点时间和焦点移动为基础来辨认事件的相对次序，是个尝试。

3. 名词短语的修饰关系。在对指称作解释时，一个复杂的名词短语的各个部分的修饰关系会产生两个问题：一个问题是对复杂名词短语所描写的客体之间的结构和功能的关系选择一个合适的基本的（符合意图的）描述，这个描述是从几个可能的描述中选出来的。例如对 the pump dispenser 的符合意图的描述取决于揭示两个物体的功能关系：pump［泵］和 dispenser［分配器］。另一个问题是设法识别所指的客体。有些带有介词修饰语的复杂的名词短语，它的每一部分在某个语境里可能有几个指称对象，识别所指的客体会较为困难，例如在一个有两只猫和两顶帽子，而又只有一只猫在其中的一顶帽子里的语境里说 the cat in the hat，虽然指称对象是清楚的，但是没有哪一种简单的方法可以作为识别指称对象的基础。在这方面的研究主要是探索一些使用语言知识来预测名词短语的各种可能结构的框架。从计算机的算法的角度看，既要考虑表示决定意图所需的特定领域知识，也要考虑规定寻找那些知识的过程。有些识别指称对象的研究（主要是那些带有介词短语的结构，而不是那些名词性复合词）则提出了所谓“递增指称评价”(incremental reference evaluation) 的概念 (Mellish 1982)，它假定中心语、内嵌在介词短语里的名词短语和介词所规定的关系对选择候补的指称对象构成一些制约，随着这些制约逐步得到满足，指称评价就会逐渐增加，一直到最后识别意图的指称对象。

4. 语句省略。语句省略指的是从语句中略去一个句法上需要的短语，而其内容仍能从前面语句中恢复过来，以决定省略语句的意义。对语句省略的合适处理是要求把语篇（而不是

句子）看成是交际的主要单位。语句省略又分两类：一类是省略的内容可以直接从前面语句的意义表征中恢复。另一类是省略的内容可根据语篇的意图结构来恢复；在这种情况下，略去的东西不一定出现在前面的语句里。

对第一类语句省略的研究主要是从名词短语和动词短语两个方面来进行。它们把省略看成是语篇中的前指式，因为省略也可以解释为到前面提到的短语及其语篇表征中去找寻恢复材料的源泉。和语篇前指式不同的是，省略的短语所指动作或客体不一定和前面提到的完全相同的，例如，

Fred kissed his mother.

John did 0 too.

0＝John 亲吻他自己的母亲，但也可以指 John 亲吻 Fred 的母亲。

在这里，提供资料来建立略去短语的表征的是前面提到的短语所显示的语篇表征。

第二种类型使用意图结构来恢复略去的材料。我们假定被略去的短语对语篇或语篇片段的整体目的是有所贡献的。只要语篇环境包括足够的信息，往往不需要使用一个完整的语句；一个短语或句子的一部分就足够了。例如，

（对火车站问讯处的服务员说）上烟台的火车？

省略式的短语就能直接体现语篇的意图，无须把整个句子恢复就能把意思传递出去。这些短语可以直接和一件言语行为相联系，而该言语行为又能反过来实现说话人在特定领域内的行动的部分的计划。

9.4.4.5. 计划的识别

对识别语篇计划的理解主要是来自哲学家 Grice 的两个论点：一个是非自然意思（non-natural meaning）的理论；一个是蕴涵（implicature）的理论。在第一种理论里，说话人使用语言时的意思主要是依赖于他们想做什么。蕴涵理论则来自这样的观察：使用语言时所传递的很多东西都不是明白地说出来的。这种理论表明，说话人所传递的东西比他们在话语里明白地表示出来的要多得多。在两种理论里，说话人在说话时有意让听话人懂得其话语所要传递的意图。

所以语言的理解要求听话人了解说话人有哪些意图：说话人想通过话语来部分实行什么计划。会话中的计划是可以识别的，听话人只有了解说话人的计划，说话人的交际意图才能实现。说话人必须在语篇中包括足够的信息，使听话人有可能识别他的计划。

在交互作用的会话里，识别计划有助于回应说话人的语句，例如一个人到车站服务台打听去天津的火车什么时候离站，“您知道往天津去的下一班车什么时候出发吗？”说话人的意图是提出一个询问。如果服务员识别他的计划，就会完成某些动作，如说“下午两点。”而不是回答“对！”有时交际计划是更大的计划的一部分，听话人必须推断出这个更大的计划，例如说话人的计划是去天津办事的大计划的一个组成部分，他就能提供更多的信息，帮助乘客达到目标，如告诉他从那一个门进站，甚至告诉他，“两点钟出发的那一次车是慢车，两点半还有一班直达快车。”

在自然语言处理中，人们试图采用一些逻辑的方法来表示计划的识别，70 年代的 Fikes & Nilsson（1971）首先研制了 STRIPS 系统。这个系统虽然比较粗糙，但可见一斑。一个计划可以看成是一系列的动作和状态，具体说，$[\alpha_1, S_1, \cdots, S_{n-1}, \alpha_n]$ 是一个把状态 S_0转变为状

态 S_n 的计划，只要

●α_i所有前提在 S_0 是真的；

●α_n 的所有效应在 S_n 是真的；

●在每一个中介状态 S_i 里，α_i 所有的效应和 α_i+1 所有的前提都是真的。

形成计划就是把目前的状态转化为另一个其中某些目标是真的状态。Allen（1979）还进一步提出一些计划识别规则，其中一条是动作—效应规则（Action-Effect Rule）：

Bel（H，int（S，α））⇒Bel（H，Int（S，e））如果 e 是 α 的效应

这条规则的意思是，"如果听话人（H）相信（Bel）说话人（S）的意图是完成某个动作 α，那么听话人可认为，如果 e 是 α 的效应，说话人的意图是使某个命题 e 成为真的。"意向性是识别计划过程的先决条件，这就是说，已知 H 看见 S 在完成某个动作 α，只有在 H 认为 S 是在有意做 α 时，动作一效应规则才能起作用。

这种表示法的缺点是它只能适用于单一的语句，而在典型的语篇里，几个语句是按时间次序展开的。于是又有人（Carberry，1988）提出一种递增识别技巧（incremental recognition technique）。在递增识别里，计算机系统扮演了 H 的角色，开始时使用上述基本技巧，尽量从 S 的话语中进行推断。这时，系统仍未能识别 S 正在执行的几个计划，但是这些中间结果可以成为部分计划的候选。然后系统在处理后面的语句时，再把这些部分计划加以扩充。如果我们从说话人的注意焦点方面来考虑，递增计划识别就会变得更为突出。

计划识别过程会因为说话人同时执行几个有关系的计划，或几个计划交错在一起，而变得更为复杂。对语篇中的计划识别的研究目前正处于一个方兴未艾的阶段。一般的倾向是联系 Austin 和 Searle 的言语行为理论来深入研究，这样一来，交际行为就可以看成一种有意图的活动；而且用来分析非言语活动的那些工具同样也可以用来分析言语活动，例如在人工智能方面，计划识别的模型就可以利用控制论的自动化形成计划的模型。

10. 实验设计

上面仅是一般地介绍一些实验方法，特别是认知和心理语言学的实验方法；应用语言学的一些基础研究也采取这些方法，但是应用语言学的应用研究的方法尚未接触到。这些实验也还有一个设计方法的问题。往往出现的情况是实验方法的选择是可行的，但是却缺乏科学的设计方法支持，以致所收集的数据不可信或无效，或是缺乏说明力和推断基础。应该指出的是，设计一个实验，既要有科学性，要求有严格的规程，反对任意性；但同时也要有创造性，要求注入新思想，反对形式化。实验都必须精心设计，使我们有所发现，有所前进。

10.1. 选择课题

研究人员有时不知道从何入手来开展一项新的研究，有时选择了一些琐碎的或不切合实际的课题，有时所选择的题目和所要说明的问题不能呼应。这都说明，选择课题对整个研究起到关键的作用，可以说是研究的预备阶段。

一个课题有如下的特征：

1. 课题应就两个或更多的变量之间的关系提出问题，例如性别是一个变量，外语学习成绩是另一个变量，如果我们认为这两个变量间存在一定的关系，女学生的外语学习成绩会比男学生的要高，我们就可以形成一个研究课题。但是我们又觉得这样看问题会过于简单化，学习成绩好坏和学习阶段也有关系，于是我们多引入一个变量，低年级的女学生外语学习成绩会比男学生的要高。

2. 课题应该明确而毫不含混地，并通常以问题的方式提出，例如，学生的性别、学习阶段和外语学习成绩有些什么关系？

3. 课题应该可以用实验方法（即收集数据的方法）来检验的，换句话说，这种实验是可重复的，课题研究所产生的结论，别人也可以通过实验而取得。

研究课题源于生活，是现实的需要。但我们可以把这个预备阶段分为不同的渐进的阶段，也可以说这些阶段是循环式的上升。Seliger & Shohamy（1989）提出四个阶段：（1）一般性课题；(2)课题的集中；(3)决定目标；(4)形成研究计划和假设。并用图形来表示，如图10.1。

10.1.1. 一般性的课题

一般性课题可有不同的来源：

1. 个人的经验和兴趣。从个人的语言学习和教学中常会涌现一些饶有兴趣的课题，对外语教师来说，第一语言和第二语言之间的关系是一个永恒的主题，有的外语教师还养成一种写教学日记的方式来记录自己的观察；或鼓励自己的学生写学习日记，记录自己的学习和使用外语的经验，甚至自己的有意识的思维活动，这些观察虽然带有主观的个人成分，但往往

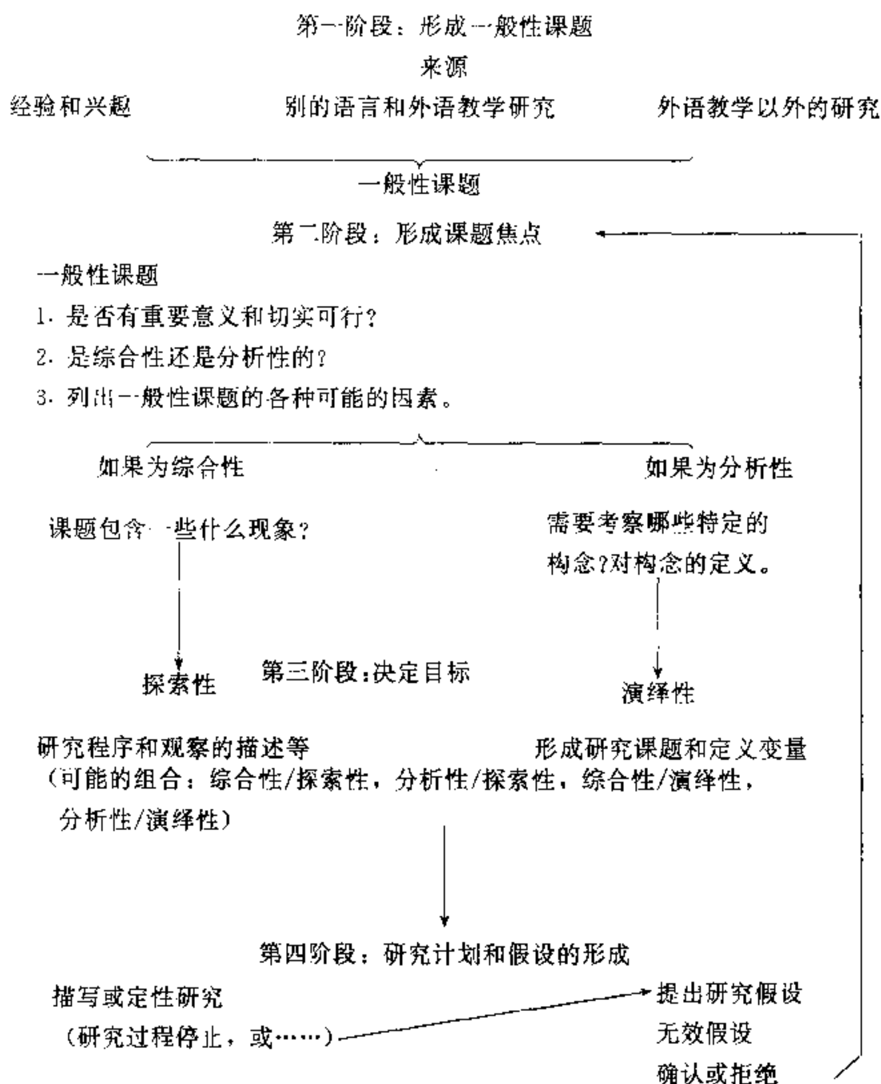


图 10.1

可以作为实验设计的出发点。

2. 阅读其他关于语言和外语教学研究的文献。这不但是为了了解这些领域的最新发展，而且也可启发我们通过自身的实践对一些新的理论、方法的主张作进一步的探究。对我们有启发作用的研究有两种：

(a) 理论性研究。这些研究或者提出一种理论，或者对其他理论作出综合，或者讨论某一种理论（不一定是外语教学本身的理论）对外语教学的启发，都能开通思路，引导我们去作实验性的研究。例如在理论语言学中语言普遍现象的理论，言语行为（speech act）的理论不但在语言学界引起很多争论，而且对外语教学的理论和实践都发生很大影响。这些理论研究都有一种生成假设（hypothesis generating）的作用。

(b) 实验性研究。这些研究可以是探索性的，或演绎性的；也可以是以某一种理论或假设为基础的，或不以之为基础的。但都收集了一些关于外语教学的数据，这些结果往往有启迪的作用，因为我们都可以用自身的实践，并进一步设计自己的实验去验证和质疑。就语言

普遍现象理论而言，如果用管辖和约束理论来解释第二语言习得，那就应该用它来预测学生会犯些什么语言错误，预测语法规则习得的次序，预测第一语言中有哪些规则会转移到第二语言习得，哪些规则必须独立学会等等。

3. 外语教学以外的研究。现代语言理论研究的一大特点是它的多元化发展，它与社会学、心理学、神经生理学、教育学等相结合产生了一系列的边缘学科。这些研究的成果促进了外语教学的发展，也给外语教师提出很多饶有兴趣的课题。社会语言学对语体的研究促使了外语教学大纲的改革，乃至专门用途外语的出现；心理语言学在认知和语言，心理和语言方面的研究，使外语教师深入了解他们教学对象的学习心理过程，对怎样培养他们的语言能力富有启发意义。例如对阅读过程的研究不但大大促进了阅读能力的培养，而且开拓了很多新研究领域。计算机的普及使计算机外语辅助教学、计算机辅助外语测试蓬勃发展，大有一日千里之势。外语教学以外的研究不但给外语教学注入了新思想，而且也向外语教师提出了挑战，他们必须掌握新的理论和方法，去解决许多新的课题。

10.1.2. 课题焦点

10.1.2.1. 可行性问题

一般性课题被决定以后就必须充分考虑这个课题是否可行，以免仓促上马，而又中途改变研究计划，造成资源浪费。下面联系外语教学研究提出几个需要考虑的可行性的问题：

1. 怎样找到一般性课题的答案？要承担什么？是否需要作实验才能取得答案？是否需要测试或问卷？要解决这个问题有几种可能性：既可在学校环境，也可在自然环境里研究被试的学习外语的过程；既可通过观察和描述来进行综合性的研究，也可集中在外语学习的某些方面（如句法形式或语篇策略）来进行分析性的研究。我们也可以找一些已知的不同水平的学生，来检测他们的哪些特征与学习好坏有关。总之，研究的方法是多样的，但各有优缺点，我们应该从一开始就考虑使用什么方法来取得答案。

2. 研究者是否具备考察该问题的先决的背景知识？是否需要有语言学和社会语言学的知识？需要牵涉到多少统计学知识？是否需要更专门的人参与？如果研究牵涉到一些周边领域，我们需要作多少研究才能继续进入另一个阶段？对一个课题的研究可有不同的方向，方向不同，所要求的先决条件就各异。如果目标是观察外语学习的社会过程，我们就需要具备诸如群体行为、群体语言交际型式、社会环境对外语学习的作用等等社会语言学知识。如果研究的焦点为学习好坏不同学生的语言能力所牵涉到的语言因素，我们就需要多懂得一些语言理论、收集和分析语言数据的方法。

3. 一般性课题所使用的术语和概念是否明确定义，而且前后一致？它们是否和别的研究者所使用的一致？例如不同人对“语言学习”(language learning) 和“语言习得”(language acquisition) 的定义会有差别，我们取哪一种说法，应该从一开始就很明确。就习得而言，也还有一个定义的问题，是按测试的分数去定义习得呢？还是从功能的习得（如能够执行“提出请求”的语篇功能）方面去定义呢？这并不是说这些术语的定义必须是大家都接受的，但是如果有明确的操作定义 (operational definitions)，就不会引起含混和误解。

4. 预期会出现什么有关后勤的实际问题？研究一般性课题牵涉到是否有足够数量的被

试？需要多少时间才能完成这项研究？被试能否坚持到底？需要一些什么仪器，如录音机、录象机、耳机、计算机等等？谁来收集数据？是否需要训练助手？如果需要计算机分析，有没有计算机时间或别人来帮助？研究者是否需要学会使用一些计算机统计软件包，如 SPSS、SAS、STATISTICA 等等？

10.1.2.2. 综合性还是分析性研究？

接着可行性问题以后需要考虑的是在综合性或分析性研究中选择较为合适的一种方法。在 6.1.4 里，我们已对综合法作过介绍，综合法要求全面的研究，因为所研究的因素互相联系，难已分离。分析法则选择其中某一个或几个因素，进行深入的分析，一般需要把那些不准备进行研究的因素控制起来。例如我们选择了一般性的课题，“为什么学习外语的学生有不同的学习进度？”这个课题中的一些因素的考察宜于用综合法，有些又宜于用分析法。但是也有一些因素是两种方法都可使用的，如：

- 学习者以前的学习经验
- 学习者对待学习语言的班级、教师或教材的态度
- 学习者的语言学能
- 学习者的母语
- 学习者的性别
- 学习者在课内和课外的练习量
- 学习者的学习经验—课堂操练和外语的交际使用的比较
- 学习者的性格
- 学习者的认知特征

我们可以根据需要而决定采取哪一种方法。有些因素需要从全面考察，就应用综合法，例如我们觉得学习的效率和课堂上的练习的方式和分量有关，就应该从课堂练习的各个方面去观察：语言型式的练习、小组练习、个人练习、有控制的练习、自发的练习、两人配对练习等等。只观察某一种练习就容易歪曲练习的作用。采用综合法可以使我们评估各种练习形式的相对的作用。我们所关注的是练习的“综合的”概念，因为我们并不知道哪一种练习和多少分量的练习对学习效率有影响，我们只好观察各种各样练习，看它们是怎样交互起作用的。

我们也可以采取分析法来深入观察某一个成分，这通常意味着把其他的成分加以控制。例如在研究练习的作用时，我们感觉到个人练习的某一方面（如在课堂里的操练）对学习效率有影响，我们就专门观察这一方面，而把其他方面稳定下来。这就是实验的方法，可以看到实验方法使课题的研究更为集中，更容易形成焦点。

10.1.2.3. 缩小课题范围

一般来说，可研究的课题很多，而且前人都有所接触，如果我们单就两个变量的关系去提出课题，课题的范围都会过于广泛，难以操作。这就必须逐步缩小范围。一种做法是把研究领域作一个单维的分类，如图 10.2。

这是外语教学研究中的一个简单的分类，我们要选择课题的话，首先是看自己对哪一个方面最感兴趣，最有能力去进行研究。例如选择学生因素，然后再把这个范畴分为一些次要的范围，如年龄、性别、智力、学能、学习动机、兴趣、学习策略等等，再选择其中的一个

方面，然后提出自己感兴趣的课题。Tuckman（1978）提出一个三维的选择课题模型可供参考，这三维是：（1）所具备的输入；（2）教学活动和组织；（3）期望的结果。

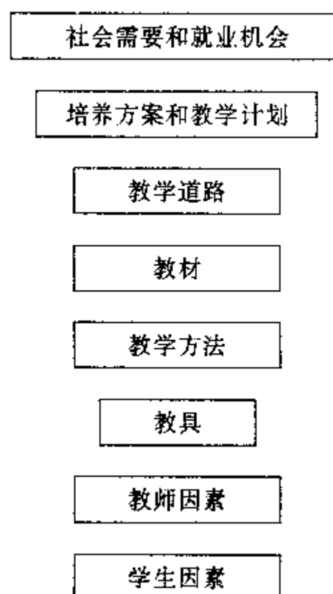


图 10.2 单维的分类

下面是作者根据这三维来显示外语教学的选题框架：

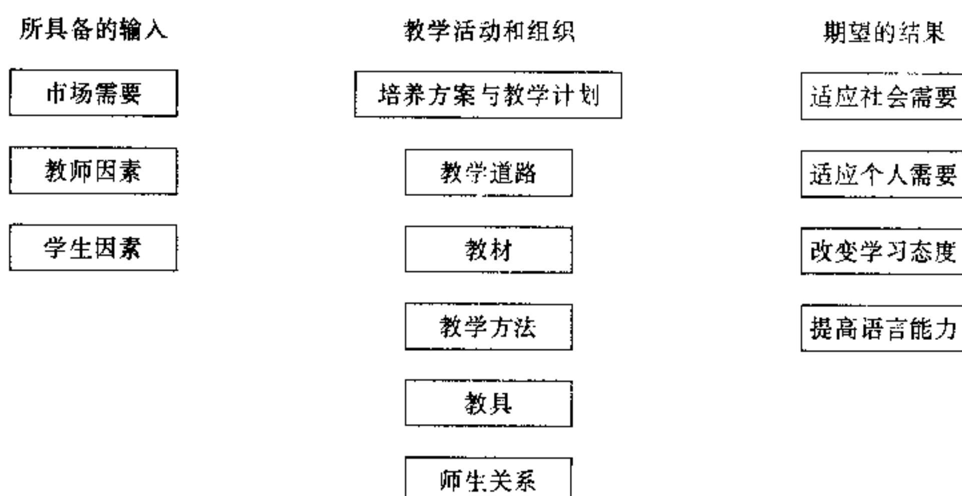


图 10.3 多维的分类

使用这个框架必须先在第一列中找出自己感兴趣的那一个范畴和次范畴，然后再到第二列和第三列中找出自己感兴趣的范畴和次范畴。例如把学生和教学方法和提高语言能力联系起来，然后再从中选择更具体的次范围。选择范围时，不一定要从一到三，也不一定三维都需要考虑。例如把就业机会和改变学习态度联系起来，把教师和教具联系起来，加以考察两者的关系。

10.1.3. 决定目标

方法决定以后，就需要决定目标。研究的目标是发现和描述呢？还是检验一个以前人研究为出发点的假设呢？如果目标是前者，那么我们的研究是探索性的（heuristic）；如果是后者，那就是演绎性的。探索性研究可以是综合性的，也可以是分析性的。综合性/探索性研究较为普遍，因为既然是探索性的，那就应该不带任何框框，全面地考察各种因素。如果我们觉得某一种因素在课堂教学中起作用，但是我们不知道是什么作用，我们也没有什么理论和假设，那就使用分析性/探索性的方法。探索性研究采用定性的方法，读者可参看第六章。

演绎性研究牵涉到提出假设和预测，并对它进行验证。假设通常是以对所研究的行为作出解释的某一理论为基础。演绎性研究可以和综合性或分析性研究相结合，构成综合性/演绎性研究（例如考察一大群变量和语言能力的关系）和分析性/演绎性研究（例如其中一个变量和语言能力的关系）。这两者可以互为补充。Purcell & Suter（1980）首先考察12个和正确发音有关系的变量，然后通过统计程序逐渐把变量数目减少至最能预测发音准确性的两个因素。由此可见，这个缩小范围的过程也是形成课题焦点的过程。

10.1.4. 形成研究计划或假设

如果研究是探索性的，这个最后阶段就是决定合适的程序和设计收集数据的方法。如果研究是演绎性的，程序就比较复杂，因为它需要明确地表示变量之间的关系，提出假设和预测。要"证明"一个假设比较困难，一般是提出一个无效的假设，假定两个变量是没有差异的，然后再通过实验数据来看无效假设应该接受还是拒绝。拒绝了无效假设，我们的研究假设就能成立。

10.2. 提出假设

10.2.1. 什么是假设？

上面说过，课题是就两个变量的关系提出问题，假设就是试图去回答这个问题，它有如下的特征：

1. 它对两个变量的关系提出一种设想，例如勤奋和学习成绩有正相关的关系，年龄和外语学习没有绝对的正相关的关系（即年纪越小，学习外语越好），学习策略对语言能力的提高有重要的影响，语言实验室对外语学习有（或没有）明显的作用等等。

2. 它必须用陈述句的形式把这两个变量的关系明确地、毫不含混地表示出来，例如"学习越勤奋的学生，其学习成绩越好。""语音实验室对外语学习没有明显的作用。"

3. 它应该是可以检验的，即假设可以重新表示为一些可操作的形式，而这些形式又是可以在数据的基础上评估的。

10.2.2. 观察和假设的关系

不能把假设和观察混为一谈。观察指所看见的东西，如去听一堂英语课看见学生积极参加教师所组织的课堂活动；根据这些观察，我们可以推断学生学习英语的兴趣很浓。我们本来是不清楚学生是否有兴趣学习英语的，我们仅是作出一种推断，这就是假设。这个假设牵涉到两个变量：积极参加课堂活动和学习兴趣之间的关系。这个假设可称为特殊假设，或预测（prediction）。为了证实这个预测，我们就去听一些别的科目的课程，或到别的学校去听英语课，发现学生参加课堂活动的积极性很不一样。我们还可以找一些学生谈，了解他们对哪些科目感兴趣等等。最后我们就形成一个一般性假设：到课室去听课可以了解到学生的学习兴趣和学习积极性的高低。这个一般性假设的概括性更高，它必须通过更多的观察和检验预测来证实。当然我们不能到所有的教室去听所有科目的课，只能抽样。根据抽样而作出的结论就有一个概率的问题，即我们的假设在多大的程度上是真实的。检验特殊假设比一般性假设需要少一些观察，所以从检验的角度看，我们往往把一个一般性假设重新设定为特殊假设。

假设可以定义为根据我们对假定的变量关系的概括而作出的事件期望。假设是抽象的，与理论和概念有关；而观察则是用来检验特殊的、以事实为基础的假设。这里所说的观察是广义的，包括定性的、定量的和实验方法的手段。

10.2.3. 假设是怎样来的？

如果我们对一个课题作出陈述："A 和 B 是什么关系？"有三种可能的假设：

(a) 有关系。A 增加，B 也增加。
(b) 有关系。A 增加，但 B 减少。
(c) 没有关系。A 和 B 没有关系。

这是最基本的线形关系。也有更为复杂一点的关系，如开始时 A 增加，B 也增加，但到后来即使是 A 增加，B 也不再增加。如果同时考虑更多的变量，可能的假设还会大大增加。

我们决定了要研究的变量关系以后，就可采取两种逻辑手段去提出一个假设。一种是演绎法，从一般到特殊。例如我们常说一个人在某一种活动中的能力越高，他在活动中所花的时间就越短。这是我们通过演绎来作的推导：他花的时间少，是因为他的工作效率提高了。但是我们也可以演绎出另一个假设：一个人在某一活动中的能力越高，他在这个活动中所花的时间就越长，因为他越来越喜欢这个活动，而避免作哪些他能力低的活动。由此可见，我们所演绎出的特殊假设取决于我们的理论的、一般的取向。我们从一般性期望到特殊期望的过程就是演绎的过程。另一种是归纳法，从特殊到一般。我们从特殊观察出发，通过各种观察最后归纳出一个一般性变量关系的陈述。有的人通过查阅有关文献，根据他人的发现而归纳出一个假设。有的人通过探索性的研究而最后产生一个假定的陈述。

10.2.4. 建立备择假设

一般来说，我们可以从一个课题中推导出不只一个假设。例如我们要考察学生的性格和教学程序这两个因素对学习成就发生什么作用？可以有三种不同的假设：

1. 组织得较严密的教学程序使具体思维的学生中取得较大的学习成就；组织得较自由的教学程序使抽象思维的学生中取得较大的学习成就。

2. 组织得较自由的教学程序使具体思维的学生中取得较大的学习成就；组织得较严密的教学程序使抽象思维的学生中取得较大的学习成就。

3. 组织得较严密的和组织得较自由的教学程序使具体思维的和抽象思维的学生中取得相同的学习成就。

这些都是备择假设（alternative hypotheses），三个假设中只有一个符合实际。换句话说，证实了一个就自动拒绝其他两个。选择哪一个假设要使用演绎和归纳的手段。有的研究说明，当教学方法和学生的性格保持一致时，学生就能够从经验中多学到一些东西。具体思维的学生和组织严密的教学程序较一致，而抽象思维的学生组织自由的教学程序较一致，因此逻辑的推导是第一个假设。Tuckmen（1978）的实验也支持这个假设。

我们使用演绎和归纳的手段来形成假设，这意味着对有关的理论和以前的研究成果都必须给予充分考虑。研究的一个目的是充实能够对实际问题提供答案的理论，所以要尽量作出理论的概括。建立假设和检验假设可以概括我们的发现，避免就事论事。

10.2.5. 在概念化基础上的假设

研究现实的问题可以从两个层面上进行：一个是具体的可操作的层面，从可观察的角度来看事件；另一个是概念化的层面，从事件和其他事件的基本相同性（如因果关系）的角度来看事件。概念化的层面使我们对个别事例进行抽象，更深刻地理解事物的本质。假设的形成往往需要从具体层面上升到抽象层面，使我们的结论更具广泛性。例如我们可以比较程序教学[①] 和传统教学。这两种方法都是可操作的，但是为了使比较取得更高的概括性，我们就必须从教学思想和原则等方面去考察它们的异同。从概念上比较可操作的程序的异同的过程就是概念化（conceptualizing）或维度化（dimensionalizing）的过程。所以这两种方法可以从下面几个维度来加以比较：反馈程度、巩固程度、教学方式、进度控制、教学分量、学习行为。这六个维度或概念可以用来区分任何教学模型。Tuckman 还举了一个虚拟的例子：一个教育部门决定举办三种在职的教师训练班，以培养教师帮助后进学生。表面看来，这项研究是比较三种训练方式，看哪一种方式效果更好。但是比较不一定能指出效果好的训练班有些什么特征，使它比别的训练班更成功。所以与其比较哪一种方式更为成功，还不如从不同的维度或概念来比较：第一个维度是训练班的组织严密程度：

① Programed Instruction，曾经一度很流行的教学方法，其理论基础是行为主义心理学。在语言实验室所做的语言练习就是一种程序教学。

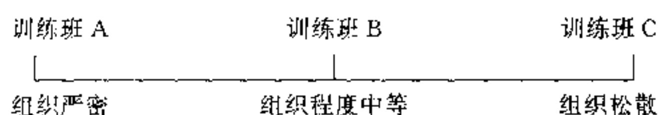


图 10.4　按组织严密程度比较三种训练方式

第二个维度是训练班所研究的问题。一种训练班强调认知方面的问题（如问题求解），另一种训练班强调感情方面的问题（如情感和态度），见图 10.5。

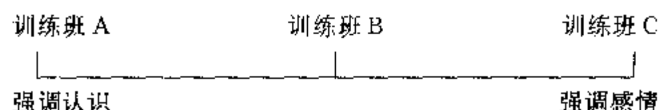


图 10.5　按所研究的问题比较三种训练方式

训练班 A 用传统方法组织，结构严密，而且强调认知；训练班 C 则相反，组织较松散，强调人际关系。训练班 B 则间乎两者之间，既考虑“理智”，也考虑“感情”，因此假设是它的效果应该最好。由此可见，在概念基础上形成的假设比操作层面上的假设更为深刻。

10.2.6. 检验假设

检验假设的目的是决定它受事实支持的程度。假设就两个变量的关系提出一种设想，要检验这个设想是很困难的。例如我们说“所有的天鹅都是白色的”，要检验这个假设就必须考察世界上所有的天鹅，这当然不很实际。我们一般的做法是使用反证的方法，即检验其反面：“并非所有的天鹅都是白的”，只要在样本中发现一只天鹅是别的颜色的，上述的假设就能推翻。如果没有发现，就接受该假设。这种反面的、“无差别的”假设叫做 null hypothesis，汉语称为无效假设、零假设、虚假设，或索性称为无差别假设。无效假设认为两个变量之间并没有真正差异，即使是有一些细微差异，这些差异可能偶然的机会造成的，不能算是真正的差异。如果经过实验发现两个变量之间的差异很大，不可能是偶然机会造成的，我们就拒绝无效假设，而接受其相反的备择的假设。

例如我们要比较两种教学方法的优劣，有三种可能的假设：

(a) 教学方法 A 比教学方法 B 更为有效。
(b) 教学方法 B 比教学方法 A 更为有效。
(c) 两种方法都同样有效，或两种方法都同样无效。

(c) 就是无效或无差别假设。当研究者提出教学方法 A 比教学方法 B 更为有效的假设时，从统计上检验的角度看，这个假设就会隐含着一个无效假设：两种方法都同样有效。如果实验结果说明方法 A 比方法 B 有很大的差别，而且并非偶然原因所造成，无效假设就可拒绝，而接受备择假设。一般来说，我们不应说方法 A 比方法 B 更为有效的假设得到“证明”，拒绝无效假设意味着方法 A 在很大程度上和方法 B 是有差别的。我们不应说一个假设是绝对的真实或错误的，因为存在着各种误差，使我们接受一个错误的假设或拒绝一个正确的假设。

我们当然没有必要以无效的形式去提出一个假设。用肯定的方式来提出和讨论一个假设

更为容易。但是为了统计上检验的方便，我们常使用无效假设。

10.3. 文献评论

10.3.1. 评论的目的

任何认真的研究都应该包括有关的文献评论，这是研究的一个有机的组成部分。我们在研究的开始针对互相联系的思想和概念的关系提出假设，即期望的关系，然后通过把思想和概念转化为收集数据的程序来检验这些期望。最后对在这些数据基础上所得到的结果进行解释，并延伸为新的概念。但是原来的思想和概念是怎样来的呢？它们是怎样联系起来形成假设的呢？它们在某种程度上来自研究者本人的头脑，但是在更大的程度上则来自前人的集体努力，来自文献资料。这些资料可以起几种作用：

10.3.1.1. 找出重要的变量

文献可以帮助我们了解一个研究领域内哪些变量是重要的，哪些是不很重要的。形成课题，即选择那些你感兴趣的，而又掌握了足够资料的变量是很不容易的。我们可能选择了一个自己感兴趣的领域，例如学生怎样学习外语，但是却不清楚在这个领域内有哪些变量，这些变量中又有哪些是热门的课题，哪些课题的发现最多，哪些课题碰到什么棘手的问题，哪些课题是足以影响全局的关键等等。查阅有关的文献不但可帮助我们找出变量和定义变量，而且还可以帮助我们认识这些变量之间的关系，认识变量在我们研究课题中的地位和作用。

10.3.1.2. 明确研究方向

要进行开创性的研究，就必须了解前人的研究，以避免重复。前人的工作也可以理解为后续工作的跳板，后续研究总是在前人研究的基础上延伸，所以仔细查阅前人的重要研究有助于我们明确研究方向，选择新的突破口。很多研究最后都指出今后的努力方向，对我们开展新的研究都有启发意义。由于缺乏研究前人的文献，我们的一些研究往往流为低水平的重复劳动，浪费了很多资源。

10.3.1.3. 综观全局

把过去的研究加以总结，有助于综观全局，了解现状。这样的活动能够对所研究的课题产生有用的结论，并了解其应用范围。我们所研究的领域往往已积累很多知识，浩如烟海；把它们整理和归纳出若干主要的结论，对己对人都可以成为研究的新起点。

10.3.1.4. 决定意义和关系

我们必须对命名和定义所找出的变量，并将其组合到课题和假设中去。这是一件艰巨的工作，必须把它们纳入广阔的环境里，才能显示其意义。研究领域的有关文献就是这个广阔的环境。如果每一个研究者都从头开始，对每一个变量都赋予新的意义和定义，都建立新的

变量关系，其结果就不是归纳性的，而是造成一团混乱。要进行有意义的研究，必须考察各个变量之间以前的关系，建立一个环境和进一步研究的方案。这样一个决定意义和关系的过程会帮助我们了解所研究的现象，并对读者作出解释；而且还能提出变量之间值得探索的关系。它能提供有用的定义，建议可能的假设，甚至提供怎样进行研究的思路。Rosenshine (1976) 考察了一系列对影响学生成绩的教师行为的研究后，作出一个在各个研究中较为一致的变量关系的综合报告：从综合报告中看到，学生的成绩和下列的教师和学生行为有密切关系：花在学术活动上的时间，直接的、范围窄的提问，学生注意作业，教师的赞扬。这说明教师对学生的指导是极其重要的，由此就可以生成一系列的假设，可作进一步的研究。

表 10.1　　将学生成绩和师生行为联系起来研究的结果的综合报告

变　量	Stallings Kaskowitz	Soar	Brophy Evertson	其他
直接花在学术活动上的时间	+			−
花在非课程活动上的时间	−	−	−	
花在学校和教学上的时间			o	+
讲授内容				+
直接的、范围窄的提问	+	+	+	+
高层次的、开放性的提问	−	−	−	o
学生注意作业	+	o^{+}	o	o^{+}
学生不注意作业，行为不当		−	−	−
学生编在大组里	+	+		
学生在无教师帮助下独立学习	−	−		
学生在教师指导下独立学习		+		
赞扬，成人的肯定的反馈	+	+	o^{+}	o
批评，成人的否定的反馈	+	−	o^{-}	−
接受学生的意见	+		+	
学生意见和教学有关	o	o	+	
学生意见和教学无关	−	o	o	
学生问题和教学有关	−	o	o	
学生问题和教学无关	−	o	−	
教师的主动性		− +	− o	
要求教师管理。要求维持秩序		o	−	

+ = 正面的，有显著意义的相关

o = 没有显著意义，混乱的相关

− = 反面的，有显著意义的相关

10.3.2. 文献资料来源

10.3.2.1. 图书杂志

图书杂志是文献资料的主要的来源，这些图书资料本身就包括一些文献评论。一般来说，有三类不同的图书和文章：

1. 教科书。教科书用通俗易懂的语言，简练的方式把最成熟的研究成果组织成一个体系。有的人以为教科书太简单，不可能提供什么研究素材。实际上教科书反映了一个研究领域的历史和现状，也会接触到一些悬而未决的问题，是不容忽略的一个来源。但是教科书的对象是未接触过该领域的学生，所以论述较多，篇幅较大，而组织较严密，如果不注意从头到尾通读，容易产生断章取义的毛病。

2. 综述性的文章。综述性的文章概述了一个研究领域的现状，而且在文章后面附有详细的文献书目，便于读者按图索骥，逐步深入，用处最大。例如在语言学方面，Newmeyer 组织了很多学者分头写综述，编成四卷《剑桥语言学概述》(Linguistics: The Cambridge Survey, 1988) 覆盖了语言学的各个领域，就是一例。在应用语言学方面，Kinsella 为英国信息与语言教学研究中心（CILT, Centre for Information in Language Teaching and Research）与英国文化委员会编辑的《剑桥语言教学概述》（Cambridge Language Teaching Survey, 1982, 1985)，也收录很多关于语言教学的综述。

3. 专著。专门就某一方面或几方面而写的专著，或专门发表这方面文章的杂志。这些杂志上发表的文章学术性较高，有方法，有数据，常被引用。

10.3.2.2. 电子资源

随着信息高速公路的建立，从因特网（Internet）中获取资料变得轻而易举。最早期的网络是在美国四所大学之间建立起来的，在目前布满全球的因特网中，作为学术研究中心的大学网络起到骨干的作用。在我国已建立大学之间的 Cernet（国家教委）和为公共服务的 China Net（邮电部)，都可进入因特网。学会从网上通信、查阅资料、传递文件，以获取又新又快的信息，已是刻不容缓的事。

电子资源有以下的几个方面：

1. 电子邮件。这是使用得最广泛的一种通信工具，不管是世界上哪一个角落，发信人都可以在任何时候发出信件，不管收信人是否在家都可以接受信件。国际上的电子邮件的传递非常迅速，不管路程多远，几分钟就能传到。现在电子邮件已被使用在远程教育上面，教师可通过电子邮件来改学生的作业或进行答疑。美国大学通过电子邮件来对已取得奖学金，即将来美学习的留学生传授英语。一个计划是由 South Carolina 大学组织的 EPI (English Program for Internationals)；另一个是美国大学拉丁美洲奖学金计划 LASPAU (Latin American Scholarship Program of American Universities)。教师主要是通过电子邮件教授学生写作和阅读各种专业教科书，而学生则提出各种有关文化、气候、生活方式的问题。

2. 远程登录图书数据库。我们可以通过远程登录（Telnet）的功能和世界上各大学的图书馆联网，查阅资料。这实际上是把我们的计算机成为网络的一个终端。现在已经有 350 多

个大学和公共图书馆（包括美国国会图书馆）允许用户使用它们的电子索引，查阅图书和提取文本。读者如能联网，可用远程登录到library. wustl. edu，就可以找到世界上主要图书馆的地址，然后再用ftp（文件传递协议，File Transfer Protocol）的程序，进入个别的图书馆。另外在因特网上还出现越来越多的虚拟图书馆（virtual library），这些图书馆所保存的实际上网络上的各种路径和地址，即可以通过虚拟图书馆进行在线检索，从缩略语词典到百科全书都有，查阅的目录并不限于一个图书馆。不少虚拟图书馆都是专业性的，例如美国Brown大学的语言学虚拟图书馆的地址为 http：//www. cog. brown. edu/pointers/linguistics. html，英国 London 大学的 Berkbeck 学院所建立的应用语言学虚拟图书馆的地址为 http：//alt. venus. co. uk/VL/AppLingBBK。这些虚拟图书馆收集了比传统图书馆还要多的资料，以后者为例，就有下列的子目录：

- 教学与研究机构
- 社团和组织
- 会议和讨论班
- 数据档案
- 研究生论文
- 电子杂志和邮件清单
- 作为第二语言和外语的英语
- 电子论文和出版社
- 其他资源

值得一提的是美国的ERIC（教育资源信息中心，Educational Resources Information Centre），它是美国教育部设计和美国教育学院维持的一个全国性的资源中心，包括16个资料交换中心，分布在全国各大学：有（1）成人、就业与职业教育；（2）学生咨询服务；（3）小学与儿童早期教育；（4）教育管理；（5）有缺陷和天赋儿童；（6）高等教育；（7）信息与技术；（8）社区大学；（9）语言与语言学；（10）阅读、英语与交际；（11）农村教育与小学校；（12）科学、数学与环境教育；（13）社会研究与与社会科学教育；（14）教学与教师教育；（15）测量与评估；（16）城市教育。每个资料交换中心都提供各种资料性服务。现在这些资料都可以从国际互联网获取，要想知道详情，可以在网络上如向信息技术资料交换中心查询，下面是一些重要的地址：

表 10.2　　在互联网络上与ERIC有关的地址

地　址	内　容
gopher：//ericir. syr. edu：70/11/clearinghouses	到各个资料交换中心
gopher：//ericir. syr. edu：70/11/AskERIC _ toolbox	了解怎样使用ERIC
gopher：//ericir. syr. edu：70/11/Ed	找寻其他资源
gopher：//ericir. syr. edu：70/11/cleaaringhouse/16 houses/CLL/ERIC _LL	到ERIC的语言与语言学资料交换中心
http：//www. cua. edu/www/eric _ ae	到ERIC的测量与评估资料交换中心

3. 电子论坛。电子论坛有两种：一种是 Usenet 讨论组，在国际互联网上进行；另一种是 Listserv（邮件列表），通过电子邮件进行。后者的传递速度较快。这些电子论坛分工很细，按专业和兴趣划分，可以自由参加或退出。我们可以在论坛上讨论问题，交换信息，寻找帮助，解答疑难，传递和收集资料。首先提出问题并获得很多回应的人有义务作一个简单小结。此外，在论坛上还可以介绍新书目录，发出会议通知，登招聘和招生启事，等等。参加电子论坛的手续很简单，只要按地址发出一个电子邮件，在文内打上 Subscribe（有的只需打 Sub）×××，并打上自己的姓名（名在前，姓在后），有的连姓名都不用打。参加以后，讨论组之间的信件都会作为电子邮件传递给你。下面是和语言学与应用语言学有关的几个活动较多的小组的地址：

表 10.5　　在互联网络上与语言学和应用语言学有关的电子论坛

地　　址	小组名称	内　　容
listserv @tamvm1.tamu.edu 或 listserv @ LISTSERV.NET	Linguist	语言学有关问题
listserv @ LISTSERV.NET（列表地址：AERA-D @ASUVM. INRE. ASU.EDU）	AERA-D	研究方法与测量
listserv@cunyvm.cuny.edu	TESL-L	英语作为第二语言教学
listserv@cmsa.berkeley.edu	TESLEJ-L	英语教师电子杂志

要注意的是，第一次订阅时要向以 listserv 开头的地址发信（这叫列表服务地址、listservaddress），这个地址只接受命令语句。故正文内只打小组名称；以后参加讨论时，就要用小组名称开头（这叫做列表地址，list address），正文内可打文本，这个邮件将会发送到小组的每一个人。例如，参加 linguist 小组，地址为 listserv@tamvm1.tamu.edu，正文是 subscribe linguist 名，姓。一般订阅以后会有邮件告诉你已被接纳，并有一封指导你怎样参加活动的邮件。以后通讯地址就要改为 linguist@ tamvm1. tamu.edu。TESL-L 是一个范围很大的小组，到 1996 年初，已有 94 个国家的 11，000 人订阅。参加后还可以继续参加一些分组，有：

- ●TESLCA-L　(Teaching with computers, Internet, technology, penpals)
- ●TESLFF-L　(Fluency-first and Whole Language Approaches -Online seminar)
- ●TESLIE-L　(Intensive English Programs Teaching and Administration)
- ●TESLJB-L　(Jobs and Employment Issues)
- ●TESLMW-L　(Materials Writers and Materials Writing)
- ●TESLIT-L　(Literacy, Adult Education, non-credit programs)
- ●TESP-L　(English for Specific Purposes)

邮件列表数以千计，而且经常有变化。要想了解有些什么小组可以参加，可到 listserv @cunyvm.cuny.edu，在正文内打 list global 即可。如想了解与语言和语言学有关的小组，则可到 listserv@ubvm.cc.buffalo.edu，在正文内打 GET FLTEACH FLLISTS。

4. 文件传递。文件传递可通过电子邮件、也可通过 ftp。一般文件都比较长，故需先加压缩，收到后再释放。我们也可以到国际互联网中下载电子杂志和电子书。利用这些手段来传递书稿（包括文字和图象）最为方便。

10.3.3. 文献资料检索

10.3.3.1. 选择兴趣领域和主题词

检索需要有方向和焦点，如果范围太广，其结果就不会有很大的用处。因此第一步是找出兴趣领域和主题词；一个兴趣领域本身也是很广泛的，但是如果把两三个领域一起交叉考虑，范围就能缩小。Tuckman 举了一个例子来说明：假定研究兴趣为小学教师性格和他们改变教学行为的关系，我们可以到 ERIC 的主题词“小学”去检索，但是这样的题目数以千计，很难看完。因此就需要增加一些主题词，如：改变行为、改变态度、改变策略、干预、奖励制度、心理特征等等，组合起来进行检索。这样就把范围缩小到 36 篇文献。所以关键是主题词，下面是用概念和变量两个维度的主题词编制一个兴趣领域的矩阵：

表 10.6　　按概念和变量两个维度的主题词编制的兴趣领域矩阵

	1. 哲学	2. 管理	3. 教学	4. 方法	5. 学生服务	6. 天气与环境	7. 测量与评估	8. 发展与社会	9. 师资培训	10. 经济与设施
1. 小学教育										
2. 中学教育										
3. 高等教育										
4. 职业技术教育										
5. 特殊教育										
6. 双语教育										
7. 早期儿童教育										
8. 阅读、外语、英语										
9. 数学与科学										
10. 社会研究与社会科学										
11. 艺术、音乐										
12. 体育										
13. 成人教育										
14. 国际教育										
15. 一般										

10.3.3.2. 检索有关题目和摘要

检索的可以是公开发表的文章，也可以是未发表的文章，或学位论文。到ERIC去检索很有必要，因为它收录了已发表的（标以EJ）未发表的（标以ED）文章。如果已经进入国际互联网，就可使用各种检索工具，如Infoseek，Yahoo!，Lycos，Magellan，Alta Vista，Excite，WhoWhere?，Open Text Index，The Electric Library，Deja News，Net Locator，Shareware.com来进行查询。这些软件功能齐全：既可使用多个主题词，也可使用概念来查询；检索出来的资料可包括文字、图表、图画，甚至有声资料和活动画面；资料来源包括公开发表的书籍杂志和未发表的电子论坛，还有在线的大型辞书和参考书。而且查阅的速度很快，保证都是最新的资料。

找到资料后，要及时做摘要卡片，最好是用数据库保存和管理，以便随时调用。一般的卡片应包括编号、题目、主题、作者、来源、日期、内容，等等。实验性研究还应包括实验目的和假设、实验方法、发现和结论。使用计算机的数据库具有无比的优越性：（1）可以多设主题词，易于缩小范围；（2）保存的资料可以是文字、图表、图画、照片和有声资料；（3）有的电子资料可以直接转存，无须键入；如是书本上的资料，可用光电扫描仪读入。常用的数据库有Dbase，FoxPro，Access等，Access是Windows下的微软办公室系统的组成部件，可与其他软件，如文字编辑器Word，电子表格[①] Excel，幻灯片制作器Powerpoint等互相传递文件，异常方便。如果Windows是中文版，还可输入汉字。

10.4. 决定变量

变量是一些有变化或有差异的因素。一个学生的英语能力会随着他学习英语时间的推移而发生变化，他的英语能力就是一个变量。经过一段时间后，他的英语能力和其他学生的英语能力都会有所变化，所以学生之间的英语能力就会出现差异。从实验统计的角度看，变量是一些人的特征或能力，它们随着时间而发生变化，或在不同的个体中产生差异。但是也有一些变量不会随着时间而发生变化，如性别；它们仅是个体差异，但也是实验中经常考察的变量。在外语学习中经常表示变化的变量有语言水平、学习动机、自我评估、健康等等；作为个体差异的变量有性别、国籍、母语背景、智力等等。

上面说过，假设是就两个变量的关系提出一种推测。但是一个假设所牵涉的往往不只两个变量。例如我们认为，在性别和年龄相等的学生中，语言能力的高低和授课时数有关，这在成年人中更为明显。在这样的一个假设里，有下列的一些变量：

自变量：授课时数
依变量：语言能力
调节变量：年龄
控制变量：性别
介入变量：学习

① spreadsheet，亦称为“电算表”，除Excel外，Louts 1-2-3也是很流行的。在Excel里也可使用Lotus 1-2-3的命令。

10.4.1. 自变量

自变量（independent variables）是我们为了研究它们对依变量发生什么作用，或它们和依变量有什么关系而选择的变量。我们也可以把它称为刺激（stimulus）变量或输入。为了观察自变量对依变量有些什么影响或发生多大程度的影响，我们就必须对它们进行操纵（manipulation）和测量。例如我们想知道授课时数对学习成绩的影响，就操纵周课时，一个班为 8 节，另一个班为 6 节。到学期末再观察学习成绩的变化。一个研究者在考察两个变量的关系时，会自问："如果我增加或减少 X 时，Y 会发生什么变化？" X 就是他所选择的自变量。因为他感兴趣的是，它怎样影响别的变量，而不是别的变量怎样影响它。

10.4.2. 依变量

依变量（dependent variables）是我们观察自变量变化会对它们产生什么作用的那些变量，也可称为反应（response）变量或输出。它们是有机体受到刺激后所显露的行为；随着研究者引入、移去、改变自变量，依变量也会跟着出现、消失或变化。如果一个研究者在考察两个变量的关系时自问："如果我增加或减少 X 时，Y 会发生什么变化？" Y 就是他所选择的依变量。我们称之为依变量是因为它的值是依赖于自变量的。它反映了人或环境发生变化后所产生的结果。依变量可以不只一个，例如为了观察改变周课时对学习成绩的变化，我们可以在期末组织一次考试，我们也可以进行一次问卷调查等等。

10.4.3. 自变量和依变量的关系

自变量和依变量都可以不只一个，但为了说明两者关系，我们以一个自变量和一个依变量为例。

自变量可以是离散性，表现为"有"还是"没有"，例如我们想研究使用电化教具对学习成绩的影响，我们就操纵一个班使用语言实验室，一个班不使用语言实验室。上述的周课时，也可理解为离散性的，因为两种不同的处理[①] 可区分为两个范畴。自变量也可以是连续性的，用数字来表现程度，如我们想研究听力理解和阅读理解两个变量的关系，可以比较听力测验和阅读测验的成绩，如做相关研究。在这种情况下，决定哪个变量是自变量，哪个是依变量，是有点任意性的。

自变量可称为因素（factors），而它的变化则称为水平（levels）。所以授课时数这个因素可有两个水平。在图 10.6 里，当自变量从水平 1 变为水平 2 时，依变量即随之增加或减少。

---

① treatment 指一种实验变量的安排，例如实验目的是比较两种教学方法，对一组使用甲方法，另一组使用乙方法，这就是两种不同的处理。

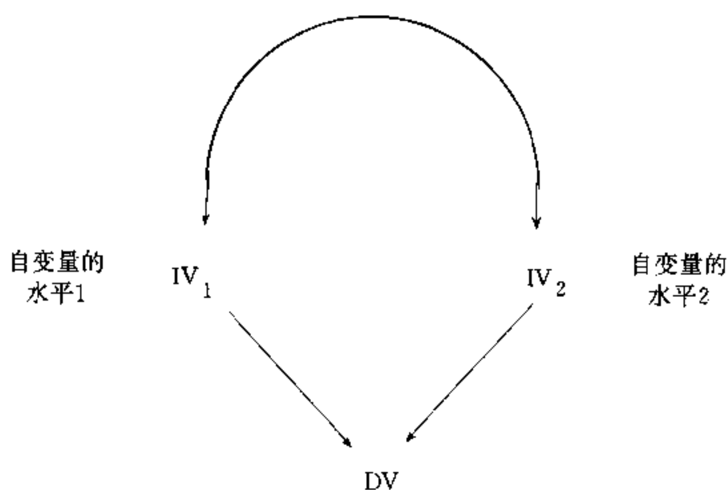


图 10.6 自变量和依变量的关系

10.4.4. 调节变量

调节变量(moderator variable)是一种特别的自变量,可称为次自变量。我们可以在实验中增加这个变量,以了解它是怎样影响或改变主自变量和依变量之间的关系。假定我们想研究自变量 X 对依变量 Y 的作用,但却怀疑 X 和 Y 之间受到第三个因素 Z 的影响,我们就可以把 Z 作为一个调节变量来分析。例如我们想观察授课时数和学习成绩的关系,但却觉得年龄(如成人和青少年)可能会调节这种关系,那么我们就可以对两个年龄组进行观察,其结果可能是无甚差异,如图 10.7a。但如果出现图 10.7b 的情况,就表明增加周课时对成人有作用,而对青少年则不明显[①]。自变量和调节变量的主要差别在于研究者怎样看这两个不同的因素,我们也可以把年龄看作自变量,而把周课时看作调节变量。如果我们把一个因素看为自变量,我们关心的是它和依变量的直接的关系;如果把它看为依变量,我们关心的是它是怎样影响自变量和依变量的关系。在教学研究中,各种因素错综复杂,应该包括一些调节因素。

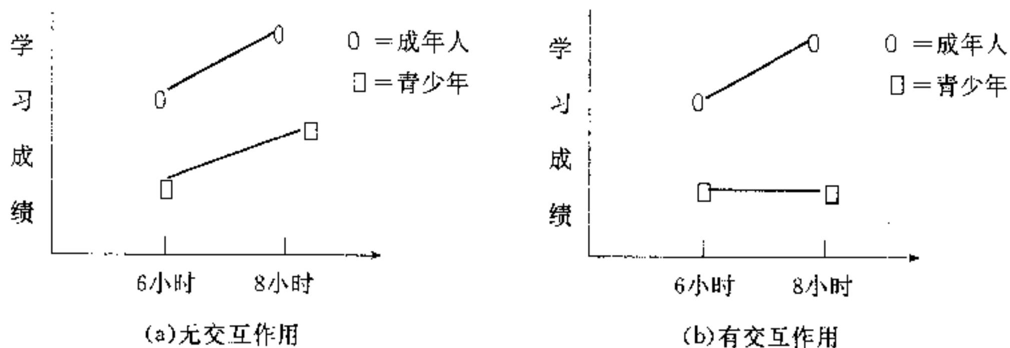


图 10.7

① 在统计分析中,这叫做交互作用(interaction effects),见 11.6。

10.4.5. 控制变量

我们不可能同时考察人和环境的所有变量。有些变量的影响必须中立化，使它们不能对自变量和依变量的关系发生作用。这些作用受到控制的变量就是控制变量（control variables）。在上例里，我们把性别加以控制，才能观察年龄这个调节变量。做法是在成人组和青少年组里，男女生各占一半。有些变量经常是控制变量，它们有时候也可以成为调节变量，如性别、年龄、社会经济地位、学习动机、智力，等等。有的控制变量和环境有关，如噪音干扰、作业先后次序、作业内容等等。有关控制组的各种问题，以后还会专门讨论。

10.4.6. 介入变量

上述的各种变量都是具体的，可以观察和操纵的。介入变量（intervening variables）则不然，它是抽象的，是一种用来解释自变量和依变量的关系的理论构建。介入变量是一个理论上影响所观察到的现象的因素；它的作用可以从自变量或调节变量对所观察的现象所产生的作用中推导出来。以授课时数和现象成绩的关系而言，介入变量可以是教学强度对学习效果的作用。介入变量是一种概念变量，它本身受到自变量或调节变量的影响，而又反过来影响依变量，故称为介入变量。

介入变量比较难于理解，但却十分重要。例如一个研究者想对比两种授课方式，闭路电视和直接讲课的优劣。他的自变量是授课方式，依变量是某种学习效果的测量。他可以自问，“在两种授课方式中什么东西使一种优于另一种？”他问的就是介入变量。一个可能的答案（只能说可能，因为介入变量既看不见，也测量不到）是注意。闭路电视也不会增加信息量，但它可能吸引听众的注意。增加注意可能提高教学效果。为什么要考虑介入变量？原因是提高概括性。如果我们认识到增加注意能提高教学效果，那就可以采取各种手段来吸引学生的注意，不一定要采用闭路电视。

我们可以再比较下面两个陈述：

1. 使用发现法（a_1）教出来的学生比使用强记法（a_2）教出来的学生更能解决实际问题（c）。

2. 使用发现法（a_1）教出来的学生会培养一种检索策略（b_1），使他能够很好地解决实际问题（c）；而使用强记法（a_2）教出来的学生会学到问题的解答，但学不到策略（b_2），因此限制他解决其他相类似的问题（c）。

在第二句陈述里，b1 和 b2 就是介入变量。只要我们在考察一个假设时问这样的问题：为什么自变量能够引起预测的结果？我们就能发现介入变量。

10.4.7. 变量的组合

为了更好地理解不同类型的变量的作用，我们不妨先看看它们之间的关系，然后再看一些实际的例子：

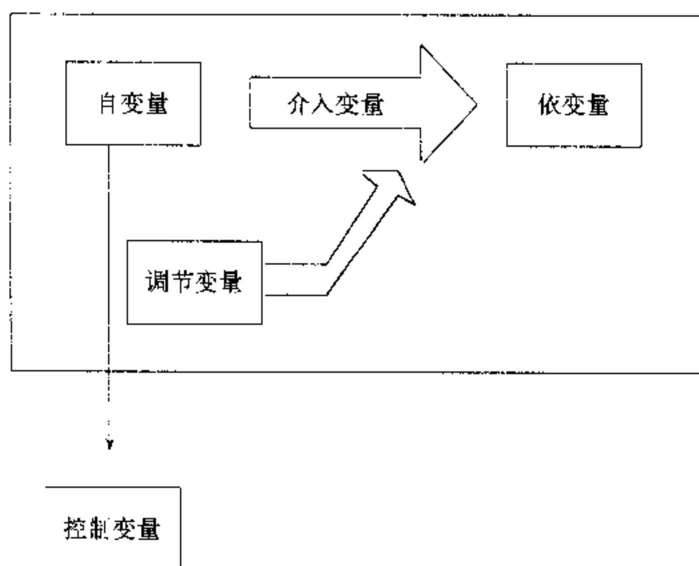


图 10.8 变量之间的关系

从图中可见，自变量和依变量的关系居于中心的位置，箭头表示研究者思考的方向，并非因果关系或时间关系。这说明整个研究的焦点指向依变量。介入变量指出连接自变量和依变量的关系或过程。另外研究者也许还想增加一个新的自变量——调节变量，以观察它在自变量和依变量之间的关系中会引起什么变化。为了保证实验不受其他因素的影响，还需要把一些因素中立化，这就是控制变量。

下面看两个实例：

1. 在 Tuckman 的指导下，Orefice 对两种性格不同的学生（抽象型和具体型）试验了四种不同的方法：（1）用录音带和手册来自学；（2）在课室使用程序教学；（3）程序教学加课堂讲授；（4）传统的课堂讲授加讨论。他的假设是抽象型的学生容易接受（1）和（2）种方法，时间花得少，效果较好；而具体型的学生则容易接受（3）和（4）种方法。在整个实验里，

- 自变量为 4 种不同水平的教学方法的比较。
- 调节变量为 2 种不同水平的性格。
- 控制变量没有在假设中提到，可能是教学内容、教学班的大小、学生的年龄、性别。阅读水平也可能是一个重要的控制变量。
- 介入变量可能为 4 种教学方式的课堂组织形式和不同类型的学生之间的关系。
- 依变量为学生的成绩、所花的时间、对教学方法的喜欢程度。

2. Brown（1988）在 UCLA 的高年级英语学生中找了两种不同类型的英语学生：一种是经过分班考试后的插班生；一种是从低班升上来的学生，再连续三个学期观察了他们的成绩（一是英语课程分数，一是系的期末考试分数，一是 50 个项目的完形填充分数）。他认为这两种类型的学生的英语水平是有差异的。在这项研究里，

- 自变量是两种不同类型的英语学生。
- 调节变量是三个学期。
- 介入变量是分班的差异。

●依变量是三种不同的成绩。

●控制变量并没有明说出来，但一些变量，如性别、母语背景、类别（本科生，还是研究生），应予以控制。

10.4.8. 变量的操作定义

对上面提到的各种变量，人们的理解和认识往往不很一致，例如什么是高水平和低水平？什么是抽象型和具体型？怎样才算喜欢或不喜欢一种教学方法？所以我们必须在实验中对所接触的变量加以具体化，给予明确的定义，这个定义往往叫做操作定义（operational definition）。操作定义以定义对象的一些可观察特征为基础。如果我们能够对一个对象或现象作出相对稳定的观察，别人也可以作出这样的观察，从而了解所定义的变量的实际内涵。例如我们把英语高水平定义为参加某一个著名的英语水平考试处于最好的10%的考生的位置，这个考试是众所周知的，高水平意味着什么，就很容易了解。进行操作定义的方式有三种：

1. 按照产生定义的现象或状态所需要完成的运作来定义。在实验中，我们往往需要使用某一程序才能使所研究的现象出现，对这一程序的描述就是操作定义。例如饥饿的操作定义是24小时内没有进食。使用这个定义，我们只要看一个人最后进食的时间，就能决定他是否饥饿。冲突可定义为两个或更多的人所处于的一种状态：大家都有相同目标，而只有一个人能够达到。食盐可定义为氯化钠。这种定义方法适合于对自变量进行定义。

2. 按照定义的物体是怎样运作来定义。例如这个物体在做什么，它有什么能动性。所以聪明的人可定义为在学校取得高分或有能力解答形式逻辑的人。食盐可以定义为可导电的水溶液。这种定义方式最适宜于用来定义教育环境中的人，因为一个人的特性表现为一些具体的、可观察到的行为。这种定义方法适合于对依变量进行定义。

3. 按照定义的物体或现象和什么东西相似来定义，例如物体的静态的特性是什么。所以聪明的学生可以定义为记忆好、词汇量大、数理能力强的人。食盐可定义为立体水晶的物质。这种定义方法既可以用来定义自变量和调节变量（当研究者不去操纵它们的时候），也可以用来定义依变量。用这种方法来定义的变量应是可以测量的。

一个假设能否进行检验关键在于我们能否对各种变量作出操作定义。在我们作出操作定义以后，就会使假设从一般到特殊，成为预测（prediction）。一个预测是一些期望的叙述，它用操作定义来取代假设中对变量的概念的叙述。所以预测可以说是一个假设的可检验的导数，是一个特殊的假设。既然变量可以有不同的操作定义，一般的假设也可以导出不同的预测（或特殊的假设）。例如我们的假设是学习的自觉性和学习效果之间存在一种正相关的关系，预测可以是一个班里学习自觉性高的学生比自觉性低的学生要学习得好些；也可以是强化训练班的学生比一般的学生的学习效果要好些；也可以是经过教师诱导而提高了学习积极性的学生比不经诱导而放任自流的学生的学习效果要好些。

到此为止，我们勾勒了从选择课题到形成假设、文献研究，再到具体的定义变量，提出预测的过程。接着下来的是怎样控制和操纵变量，设计实验。我们不妨在这里用一个承前启后的蓝图来表示这种关系：

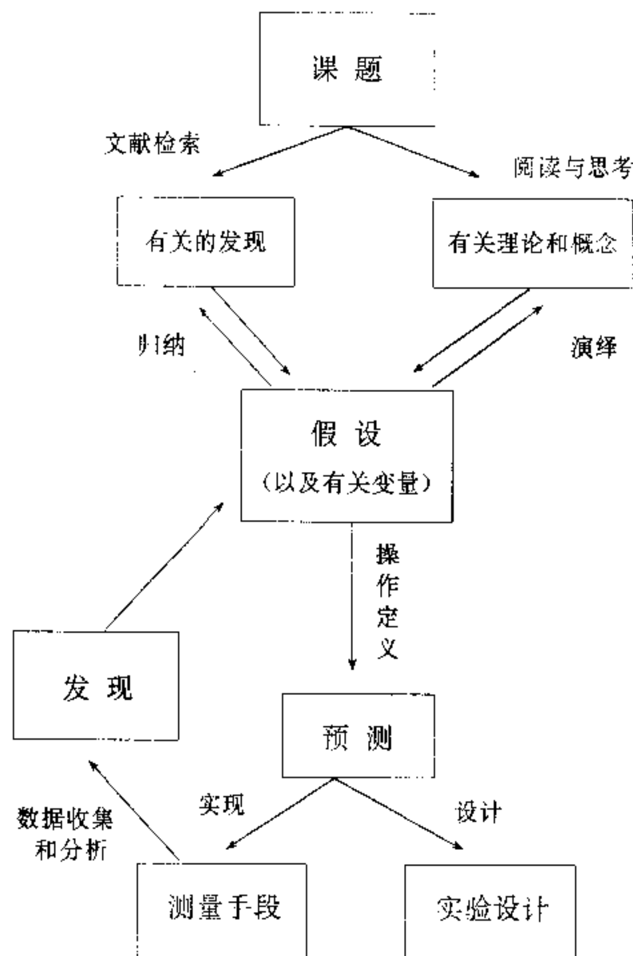


图 10.9 实验设计蓝图

10.5. 操纵和控制变量

实验性研究的本质是控制，除非对所有足以影响得到的结果的其他因素加以控制起来，我们是无从对某一种因素所产生的作用作出有效的评估。要解决这个问题，我们一般是采取控制组（或称参照组）的方法。一个控制组是一组被试，除了没有接受处理（即没有自变量）外，他们的选择和经验在各个方面都和实验组是一样的。这样才能认为依变量的变化和自变量或调节变量有关。

在表 10.8 中 Sx（水平 1）和 S0（水平 2）均为自变量，而 Rx 和 R0 均为相应的依变量，S1 到 S4 为控制变量。控制变量在两组中是一致的，如果结果 Rx 和 R0 有差异，就是 Sx 和 S0 的差异所做成的。

Townsend（1953）用一个具体的例子来说明控制组的重要性。假定有一位研究者认为他发现了一种能对低能的被试提高其智商的药物，为了检验他的药的效能，他把药作为自变量，把他的被试组的智商作为依变量。假定他在施药后发现被试的智商确有所提高，他能否认为这是药物的作用呢？他很难下这样的结论，因为他不知道是否不吃药被试的智商就不提高。也

许被试的动机在第二次检查智商时比第一次检查智商时强；也许整个实验组吃得好些，环境更好些；也许第一次检查智商的效应传递到第二次检查，使第二次检查的结果要好些。总之假定智商提高了，其原因是多种多样的。要解决这个问题，就必须使用一个控制组，组里的被试也和实验组的被试一样，同时服用一种"药"，只不过这是一种像维生素C那样的安慰药(placebo)。被试不清楚他们服用的是真正的药还是维生素C，他们处于"盲目"的状态。为了防止施药者在被试服用时会发出暗示，所以施药者也不应知道，这就是"双盲法"。建立控制组的目的是使自变量和调节变量以外的所有其他变量都保持一致。这些足以影响效度的变量可以分为两大类：影响内部效度的因素和影响外部效度的因素。

表 10.8　　实验组与控制组的设计

实验组		控制组	
环境刺激	被试反应	环境刺激	被试反应
S_x	R_x	S_0	R_0
S_1	R_1	S_1	R_1
S_2	R_2	S_2	R_2
S_3	R_3	S_3	R_3
S_4	R_4	S_4	R_4

10.5.1. 影响内部效度的因素

内部效度指的是把实验的内部因素控制起来，保证实验结果确实是我们所操纵的变量所产生的。在一个缺乏内部效度的实验里，我们往往弄不清楚组间的差异是来自所操纵的自变量，还是来自未加控制的因素。Campbell & Stanley (1963) 认为有8个足以影响内部效度的因素。

1. 历史 (History)。这是在实验研究中的一个术语，指的是在检验实验变量时，在环境中同时发生的事件。如果我们检验的是一组学生执行某一个教学大纲的结果，但与此同时，学生由于外部原因造成情绪紧张，我们所测量的结果是执行大纲还是情绪紧张所造成的，就不得而知了。要控制这个因素，就必须在实验里使用一个经历同样历史（经验）的控制组，例如实验组是在外部环境不佳的条件下执行的，控制组也应在相同的条件下执行。如果实验组和控制组都经历同样的历史，造成情绪紧张对两个组应该是一样的，结果若有差别，就不会是情绪紧张造成的。

2. 选择 (Selection)。被试的挑选对实验结果会产生影响，有的被试可能聪明一些、成熟一些或配合得好些，实验的结果就会比另一组被试好些，这可能和我们实验变量无关，由此而作出的评估就会失效。要克服这个问题，就必须保证实验组和控制组的被试都是随机安排的，即每一个被试都有同等的机会被挑选为实验组或控制组的成员。这个问题后面还会专门谈到。

3. 成熟 (Maturation)。成熟指的是参加实验的人在实验过程中所发生的变化。有的纵向研究需要较长时间的观察，在这期间会发生一些被试成长过程中的变化。这些变化，而不是

实验变量，也可能影响实验的结果。因此我们需要一个和实验组同样类型（即可能经历同样成长过程）的控制组，才能对实验变量进行观察。

4. 测验（Testing）。测验指的是在实验前所做的测验对被试在实验后所做的测验的影响。很多实验都在实验前让被试做一个测验，以决定某一个变量的起始状态；然后在实验后，再做一个（甚至是相同的）测验以观察这个变量的变化。前一个测验的经验往往能提高后一个测验的成绩，这样后一个测验就不是测量实验变量的效应。传统的解决办法是避免在实验前使用测验（它们并非那样必要的）；也有的人建议采用一些被试自己意识不到的测验办法，以减少其对内部效度的影响。

5. 手段（Instrumentation）。手段指的是在实验的测量和观察过程中所发生的变化。这些手段包括测验、机械的测量工具、观察者和评分员。虽然机械的测量工具本身不大可能会发生变化，但是观察者和评分员却很有可能在实验过程中改变他们收集和记录数据的方式。他们还会因为知道实验目的，有意无意地增加支持假设的可能性，从而影响内部效度。一个实验的测量手段和数据的收集者都必须在时间上和在组别间保持一致。

6. 统计回归（Statistical Regression）。如果我们根据一个变量的最高或最低分去选实验组，就会产生统计回归的问题。例如我们对一群学生进行智力测验，只要分数中的前 1/3 和后 1/3 来做实验，在后来的测量中就会出现高分组降低，而低分组升高的趋中的倾向。而且这两个组在后来的测验中，就是不加处理，也会出现差异。原因是偶然的因素最容易影响两端的分数，而这些偶然的因素在第二次测量中又不大可能会重新出现。因此我们必须避免选择两个极端的得分者，排除平均分的得分者。

7. 实验死亡率（Experimental Mortality）。在任何实验里，最好是从决定好参加实验的所有被试中取得数据，因为中途退出的被试可能和留下来的不一样。如果结果有差异，就会产生偏颇或出现内部失效。在一些有几个实验条件，而被试又在不同的条件里中途退出的实验里，这种偏颇更易产生。要避免这个问题，就必须有较大的样本，并对中途退出的人做跟踪处理，以保证样本的代表性。

8. 稳定性（Stability）。稳定性指的是结果只出现一次，以后不再能出现，因此它是不稳定、不可信的。这个问题可用统计的检测手段来解决，这些检测手段可以表明结果是否属于偶然事件的概率。

Tuckman 还加了两个因素：

9. 因素的交互组合（Interactive Combinations of Factors）。也有可能影响内部效度的各种因素组合在一起出现。例如选择和成熟交互组合影响效度，实验组和控制组的年龄如不加以控制，就会出现成熟的问题，因为某些年龄组的儿童成熟得比别的年龄组的儿童要快些。而且在一个年龄组里所发生的变化可能会更加系统地和实验变量有关，而别的年龄组则和实验变量的关系没有那么密切。

10. 期望（Expectancy）。往往实验组接受处理后的结果好象比控制组的好，但这事实并非如此，而是因为实验者和被试相信它更为有效，在行为上与之适应。

10.5.2. 影响外部效度的因素

外部效度指的是一项研究结果的概括性和代表性。我们在做实验研究时，总希望结果能

够在以后应用到别的地区和别的群体上面。为了使结果更具有概括性，我们必须考虑实验设计的外部效度问题。和概括性有关的因素起码有四个：

1. 测验的反作用（Reactive Effects of Testing）。如果在实验前所进行的测验对参加实验的被试产生敏感的作用，那么处理的效果就可能部分来自测验的敏化。而在另一些实验条件里，由于在实验没有使用测验，处理的效果就没有受影响。如果我们比较这两组的差异，得出前一组的结果高于后一组的结论，就会出现外部效度的问题。因为前一组的结果是测验的反作用所做成的。例如在一个关于学习态度变化的研究里开始有一个关于态度的测验，被试可能对问到的态度产生敏感，更加注意交谈，结果是态度有较多的变化。如果在别的组里，没有实验前的测验，同样的交谈就可能会产生不同的结果。

2. 选择偏颇的交互作用（Interaction Effects of Selection Bias）。如果一个研究的选样不足以代表大的总体，我们就很难根据样本的结果来推断总体。一个区域的学生参加实验所得的结果不一定能够适用于另一个区域的学生；在城市学生身上做实验所得的结论不一定能用来解释乡村的学生。所以实验的样本应该尽量代表总体。

3. 实验安排的反作用（Reactive Effects of Experimental Arrangements）。实验的安排或参加实验的经验对实验结果也会起反作用，例如两个实验者为了研究快被淹没的效应，把一个被试放到游泳池，然后往池里灌水，他们用绳子把被试系在池边以保安全。但是他们却忘了这个实验，等到他们回到池边，被试都快要被淹没了，才连忙关水。两个实验者本人给吓坏了，他们把被试拉上来，松去绳子，问那个被试，“你怕不怕?”被试安然回答道，“没事，这不过是做实验罢了。”这些反作用可以称为Hawthorne效应。它是一些研究人员在20年代调查美国西部电力公司的一间Hawthorne工厂时发现的。当时的研究是为了考察改善厂房、增加物质刺激和休息时间对提高生产率的影响。但是结果表明，不管条件如何，生产都增加了。因为工人意识到他们在参加一个实验，而且工厂领导希望得到好的结果。所以Hawthorne效应指的是由于参加实验而带来行为改善的效应。

4. 多元处理干扰（Multiple Treatment Interference）。实验的参加者同时接受几种处理，有些是实验本身的处理，有些和实验无关。这些处理产生交互作用，降低了其中的一些处理的效应。例如，让学生参加一个实验，他们还要参加一些学校的正常的活动，这些活动也是一些处理，会对学生产生这种或那种影响。光是实验处理所产生的结果往往不同于这些多元处理交互作用所产生的结果。

10.5.3. 选样的控制

控制组里被试的挑选应该尽量和实验组的保持一致，常用的方法有：

1. 随机选样。这是一种普遍使用的控制选样变量的程序，即用随机的办法把被试安排到控制组和实验组，不带有任何框框。例如我们把能够参加实验的可能的被试都写上名字（或编号），放在一起拈阄，第一个是控制组，第二个是实验组，第三个是控制组，第四个是实验组……（也可以先抽控制组的被试，再抽实验组的被试），一直到两个组所规定的数目。另一种简化的办法是查随机数字表，然后按编号分配。例如要在100人中抽50个被试，就查表上的两位数。有时我们需要在实验前进行测验，要注意贯彻随机的原则，即抽样必须独立于测验成绩来进行。有时我们用整班学生来参加实验，要注意这些班级往往不是随机分配的。

2. 配对技巧（Match-Pair Technique）。要采取配对技巧，研究人员必须首先决定哪些控制变量在实验中最容易出问题（即因为选样不当而影响到内部效度）。这些控制变量有性别、年龄、社会经济地位、民族、智力、性格、学业成绩、实验前的测验成绩，等等。然后再到可能的被试里寻找和这些控制变量等值的被试，一个11岁智力低的男童应该和另一个11岁智力低的男童配对，随机分在控制组和实验组。这样就把最足以影响内部效度的那些变量控制起来。如果在实验前有一个依变量的测验，也可以根据测验成绩来配对，例如一个关于数学教学方法的实验，在实验前先做一次测验，实验后再做一次测验，以观察不同的方法会影响学生的进步。我们也可以利用实验前的测验成绩来把被试分配到控制组和实验组，把最高分的两个随机地分到两个组，然后次之。

3. 配组技巧（Match-Group Technique）。这种方法和上述方法相同，但使用得没有那么广泛。它把个人分配到两组去，使两组在关键的控制变量上的平均值相同。例如关键的控制变量是年龄，两组的平均年龄定为11.5或11到12岁之间。两组的配对年龄往往做不到完全一样，所以我们只好要求平均值一样。但是这种方法往往会引起统计回归的问题，因此最好还是用随机的方法来选样。

4. 用被试作为自己的控制。如果所有被试既属实验组，又属控制组（如交叉分配），那么选择的变量就可以理解为得到适合的控制。但是也有一些环境的制约，使这种方法难以使用，因为实验的经验往往会对被试在控制组的行为发生影响。例如在研究教学行为的实验中，被试经过实验处理后，他在控制组的行为就会反映实验中所取得的经验，换句话说，他的成熟程度已经发生变化。如果有可能让被试作为自己的控制，那就必须使用相互平衡的方法（见10.5.4）来控制次序效应。

5. 限制总体。总体是我们要研究的“组别”，而样本是这个“组别”中被选择来参加实验的一部分人。通过限制总体的范围，我们就有可能控制那些影响实验的选择变量，例如把总体定为大学英语专业的一年级学生。当然这要付出一些代价，我们通过限制总体增加了内部效度，但却降低了外部效度。严格限制总体使我们的结论只能用来解释一些个别的群体，缺乏概括性。

6. 用调节变量来作选择手段。如果某一个别差异的统计量会对我们检验的假设产生影响，我们就可以把它作为实验中的调节变量加以操纵，从而观察它和自变量之间的交互作用。例如我们认为性别在大学低年级外语学习会有影响，我们就可以把被试分为男女生，进行分别观察和统计。这时就需要使用因子实验设计（factorial experimental design）的方法（见10.6.3）。这种设计是研究选择变量的一个重要的方法，它可以用来考察交互作用，即自变量和调节变量相互作用对依变量所产生的影响。

10.5.4. 历史的控制

除了自变量以外，控制组应该和实验组一样具有同样的经验或历史。除了实验的环境外，要保证两组都具有相同的经验是很不容易的。现实一点说，就是从同一总体中用随机的方法分别抽出两个样本：一个实验组和一个控制组。但是在实验中仍会有一些干扰因素，即由于历史而产生内部失效，应该加以控制。

1. 移去法（Method of Removal）。外部的影响应该尽量从实验组和控制组中移去，例如

噪音、干扰和环境的变化。又如在实验前或后的测试里，被试可能会问，因而产生干扰，要移去这个干扰，就要规定不能提问。上面所提到的双盲法是另一种移去主持人对被试反应产生影响的方法。

2. 一致法(Method of Constancy)。除了由于操纵自变量而产生的经验外，其他的经验在实验组和控制组之间必须是一致的。如果需要给指导语，就必须把它写下来，对两组宣读，以保持一致。实验环境也应该一样。在有些实验里，实验组需经受某一种经验，而控制组则没有此经验，就必须对时间因素、参与实验和接触材料的程度加以控制。为了保持这些因素一致，与其让被试组没有此经验，还不如让其经受一种需要同样时间和同样参与实验和接触材料程度的无关的经验。

3. 平衡法(Method of Counterbalancing)。在那些被试需要完成一件作业或测试以上的实验里，我们往往需要控制次序的效应，即控制被试作出的反应逐步转移。这些转移可能是练习或疲劳所导致的。在被试自己既属实验组，又属控制组的情况下，这种转移更容易发生。平衡法使两组都有机会先后完成两件作业。例如有两件作业A和B，被试必须每次完成一件，我们可以把每组的被试随机地分为两半。一半先A后B，另一半先B后A。

I组		II组	
A	B	B	A

这样，两个作业的相互影响就得到中和。统计方法也很简单，只要把两组的A和B的结果加起来计算即可。如果我们特别对作业的次序感兴趣，那就必须把次序作为一个调节变量。

4. 多元平衡法(Method of Multiple counterbalancing)。平衡法适用于完成两件作业两次或两次以上。但是如果被试间要完成的作业超过两件，那就必须使用多元平衡法。这种方法比较复杂，不如先看一个例子：

Lewin等人(1939)使用被试自己作控制的方法来试验三种不同的社会气氛：专制、民主、放任自流。而且被试还必须扮演不同的角色来“创造”这三种气氛，还要在气氛中当组长的领导，所以在实验中必须控制下列的因素：

(1) 经历每一种气氛的次数；

(2) 经历“创造”每一个种气氛的组长的次数；

(3) 经历每一个组长扮演每一种角色的次数；

(4) 经历三种气氛的次序；

(5) 经历三种组长的次序。

为了要控制这些因素，我们必须有六个组，每一组经历每一种气氛各一次，经历每一种组长的领导各一次，而每一个组长都扮演每一种角色两次。三种气氛有六种可能的次序，而每一组都经历一种不同的次序。每一个组长也有六种可能的次序，而每一组都经历一种不同的次序。

其实，在实验里还有些经验的组合是没有控制的，因为这需要更多的组，而且每一组需要经历更多的经验。例如组长和气氛的组合还应该要求每一组都经历每一个组长扮演适合于每一种气氛的角色，而且还有个不同次序的问题。因为组合太多，Lewin等只能把不同组合随机处理，以防止系统的偏差。

表 10.9　　用被试自己来作控制的实验设计

	第一组	第二组	第三组
时间 1	专制气氛 组长 1	专制气氛 组张 3	民主气氛 组长 2
时间 2	民主气氛 组长 2	放任自流气氛 组长 2	专制气氛 组长 1
时间 3	放任自流气氛 组长 3	民主气氛 组长 1	放任自流气氛 组长 3
	第四组	**第五组**	**第六组**
时间 1	民主气氛 组长 3	放任自流气氛 组长 2	放任自流气氛 组长 1
时间 2	放任自流气氛 组长 1	专制气氛 组长 3	民主气氛 组长 3
时间 3	专制气氛 组长 2	民主气氛 组长 1	专制气氛 组长 2

10.5.5. 选样和历史的综合控制

表 10.10 总结了控制选样（学生效应）和历史（经验效应）的程序对一项研究的内部效度（确切性）和外部效度（概括性）的影响。

表 10.10　　控制选样和历史对内部效度和外部效度的影响

	确切性	概括性
由于处理了学生效应	通过 （1）随机安排组别 （2）匹配 （3）建立统计上的等值 以控制组间在智力、以前的成绩、性别、年龄等方面的个别差异。	通过 （1）随机选择样本/母体 （2）分层抽样 使样本尽可能代表它的母体。
由于处理了经验效应	通过 （1）使用控制组 （2）对实验组和控制组提供可比的题材内容和作业 （3）平衡教师组间的效应以控制组间在经验方面的个别差异（自变量的差异除外）。	通过 （1）尽量不干预 （2）使实验和实验者的作用不明显 （3）使用“双盲法” 使实验条件尽可能代表现实生活的经验。

10.5.6. 工具的控制

被试、评分人乃至环境的变化往往会影响测量工具的读数，这是使用工具而产生的偏颇。这些偏颇可以通过下列方法来减少和控制：

1. 建立项目之间和不同时间之间的分数的信度和一致性，使测试能够统一测量一些事物。

2. 对实验前后的测试使用同一的统计量，或使用同一统计量的不同的形式。

3. 建立测试统计量的效度，使你所要测量的东西得到真正的测量。

4. 对测试建立一种相对的分数制度（常模），使分数置于同一量表。

5. 不要使用一个评分人或观察员，在整个研究过程中使用同样的评分人或观察员，并且对评判员所持的标准建立一个可转换的指数。

10.6. 决定实验设计方案

Campbell & Stanley (1963) 对实验方案的设计进行了细致的分析，为人所乐道。他们采取了一套符号系统来代表实验方案中的各种因素：

X 代表作了实验处理（treatment），而一个空格代表控制（没有经过实验处理）。如果要比较不同的处理，就写作 X_1，X_2等。

O 代表一次观察或测量，每一个 O 都有下标，以便于理解所指的是那一次观察或测量，如 O_1，O_2等。

R 代表随机化，指因素已通过随机化处理控制起来。破折号（——）代表原来的组别，表示并没有完全控制选样的偏颇。

10.6.1. 前实验设计

前实验设计表示实验并没有控制足以产生内部失效的来源，并不能称为合格的实验设计。它们之所以称为前实验设计，是因为它们只包括一些实验设计的部件。但是这种设计仍时有发现，现列于后面，以防范于未然。

10.6.1.1. 一次性个案研究

一次性个案研究（One-Shot Case Study）可以表示为：

X　　　O

在这种研究中，只对一个组进行某种处理（X），然后对组里的成员作出观察（O），以评估处理的效果。这种研究缺乏控制组（没有接受处理的组），缺乏关于经受 X 经验的人的信息，违反了实验设计的内部效度的原则。我们没有什么理由根据这种一次性研究就能下结论说 X 导致了 O。例如有某一个幼儿园试验对 5 岁儿童教英语，经过一年的教学，儿童的英语有所发展，

试验者于是下结论说，幼儿开始学习英语以 5 岁为宜。这个结论是无效的，因为这是一次性研究，并没有用不同年龄组作对照试验。我们从何得知 4 岁、6 岁、……就不如 5 岁儿童?

10.6.1.2. 一组实验前后测试设计

一组实验前后测试设计（One-Group Pretest-Posttest Design）。可以表示为：

$$O_1 \qquad X \qquad O_2$$

这种研究在实验前增加一次测试，因而对样本提供多一些信息。但是这种实验并没有控制历史、成熟、测试、统计回归等因素，因此也不能说是合格的设计。它虽然对被试的起始状态进行测量，但是它并没有处理其他的影响内部效度的因素。

10.6.1.3. 原组比较

原组比较（Intact-Group Comparison）的型式是：

$$X \qquad O_1$$

....

$$O_2$$

在这种设计中，增加了没有接受处理（X）的第二个组或控制组（空格），与接受处理的组相比较。这样一来，像历史那样的效度因素得到了控制。如果有一些偶然的因素影响了 O1，它也会照样影响 O2。但是这种设计的问题是控制组和实验组的被试并非随机选样的。两组之间的破折号说明它们是原组。而且由于被试在实验前没有进行测试，我们很难确定这两组的起点是否一样的。这种设计的问题是，它不但没有控制选样，而且也没有控制实验死亡率。换句话说，我们无从得知一个组是否在实验前的 O 就已经高于另一组，所以实验后的 O 就不可靠。

10.6.2. 真正的实验设计

上述的三种都不能说是真正的实验设计，真正的实验设计必须是对所有的内部失效进行控制。这样的设计也有几种：

10.6.2.1. 只有实验后测试的控制组设计

只有实验后测试的控制组设计（Posttest-Only Control Group Design）是广泛使用的一种真正的实验设计，可表示如下：

$$R \qquad X \qquad O_1$$

$$R \qquad \qquad O_2$$

设计使用了两个组，一个经受了处理，另一个没有经受处理，这就控制了历史和成熟。而且被试分配到哪一个组是随机安排的，这就控制了选样和实验死亡率。两组在实验前没有进行

测试，这就控制了测试效应和测试与处理的交互作用效应。这种实验控制了所有足以影响内部效度的因素，是比较理想的一种设计。

这种设计简单易行：通过随机选样可以免去实验前的测试。所以只要有可能随机化安排被试，都可使用这种设计，实验后只要比较 O_1 和 O_2 的平均分即可。

10.6.2.2. 实验前后测试的控制组设计

实验前后测试的控制组设计（Pretest-Posttest Control Group Design）可表示如下：

R	O_1	X	O_2
R	O_3		O_4

这种设计和前一种不同之处在于增加了一次实验前的测试。和前一种设计一样，它也能控制影响内部效度的许多因素，但是增加了实验前的测试，使测试效应难以控制。因此除非是我们有特别的理由要收集实验前的数据（例如想了解统计量变化的程度），最好还是用前一种为宜。实验后可以把 O_2-O_1 和 O_4-O_3平均分加以比较。我们也可以先比较 O_1 和 O_3 的平均分，如果它们是等值的，再比较实验后测试的 O_2 和 O_4 的平均分。

10.6.3. 因子设计

因子设计（Factorial Design）是在真正的实验设计基础上发展起来的一种较复杂的设计。在因子设计里，我们往往要增加一些调节变量。如果我们对上述的实验前后测试的控制组设计加以补充，多加一个调节变量（用 Y 代表，有两个水平 Y_1 和 Y_2）就可有一个因子设计：

R	O_1	X	Y_1	O_2
R	O_3		Y_1	O_4
R	O_5	X	Y_2	O_6
R	O_7		Y_2	O_8

这里有两个组接受实验处理，有两个组没有接受。接受处理的两个组又有两个水平，没有接受处理的两个组也有两个水平。如果 Y 代表的是智力因素，智力高（Y1）的那一组有一半人接受处理，而智力低的那一组也有一半人接受处理。而且哪一半接受处理，哪一组不接受，都是随机安排的。

10.4.7 的那个四种处理和两种被试的实例就可以表示为下列的因子设计图：

				自变量（X）
R	X_1	Y_1	O_1	X_1 自学
R	X_2	Y_1	O_2	X_2 程序教学
R	X_3	Y_1	O_3	X_3 程序教学加课堂讲授
R	X_0	Y_1	O_4	X_0 传统的课堂讲授加讨论

R	X_1	Y_2	O_5	调节变量（Y）
R	X_2	Y_2	O_6	Y_1 抽象型
R	X_3	Y_2	O_7	Y_2 具体型
R	X_0	Y_2	O_8	

在因子设计中，我们可以单独估计每一个自变量，也可以估计它们的联合的、交互的作用，这样就能观察到一个变量怎样受到调节变量的影响，下图表示一个有两个自变量的因子设计里各种观察（依变量）之间的关系：

	X_1	X_0	
Y_1	O_1	O_2	O_{Y1}
Y_2	O_3	O_4	O_{Y2}
	O_{X1}	O_{X0}	

我们可以通过比较 X1 下面的观察（O_{X1}）和 X_0 下面的观察（O_{X0}）来观察处理效应，也可以通过比较 Y_1 行的观察（O_{Y1}）和 Y_2 行的观察（O_{Y2}）来观察 Y 变量的水平 1 和水平 2 的效应。而且还可以比较 O_1 和 O_3，O_2和 O_4 来观察 X 和 Y 变量的交互作用。这种统计方法就是方差分析（Analysis of Variance）。

例如现有一个提高记忆力的课程（X_1），X_0 为控制变量，只有阅读而无记忆力训练。调节变量为记忆广度 Y，有两个水平，表示为记忆力强的被试和记忆力弱的被试，Y_1 和 Y_2，实验后的结果可以表示如下图：

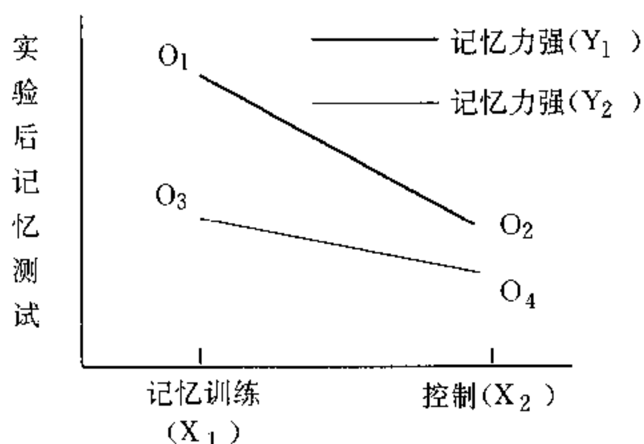


图 10.10　记忆力训练和记忆广度的交互作用

图表说明训练是有效的（实验组的 O1 和 O3 比控制组的 O2 和 O4 的成绩要高），记忆力强的人比记忆力弱的人的成绩要高（O1 高于 O3 ，O2 高于 O4），而且变量 Y 对 X 起到调节作用，

训练的效果在记忆力强的被试身上更为明显。这就是 X 和 Y 的交互作用。由此可见，因子分析的用处在于它不但使我们了解各个自变量的的作用，而且还能进一步了解它们之间的交互作用。

10.6.4. 准实验设计

准实验设计（quasi-experimental designs）是部分的真正实验设计：它能控制一些，但不是全部的内部失效的根源。虽然它们还不够格称为真正实验设计，但就效度的控制而言，它们比前实验设计要好得多。某些实验的控制条件在某种情况下难以全部实现，那就只能试用准实验设计。准实验设计在现实的条件下最大限度地实行了实验控制，使我们不至于因为各种“噪音”的影响而束手无策。

10.6.4.1. 时间次序设计

有些时候，我们由于种种原因不能在实验中包括控制组，例如整个学校系统出现了一些变革，我们就不可能找另一个（a）能够在各方面和它相比较，（b）没有进行变革，（c）而又愿意合作的学校系统。碰到这种情况，我们也许不得不使用一次性个案研究或一组实验前后测试设计，这时时间次序设计（Time Series Design）就能发挥作用。这种实验可表示如下：

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

这种设计不同于一组实验前后测试设计，它在不同的时间使用了一系列的前后测试，这就可以控制了成熟和某种程度的历史影响。时间次序设计也可以控制测试效应，因为有多次测试。当然历史的影响还不能完全排除，但它可以控制到最低限度。如果真有什么外部因素的影响，它就会从 O_1 到 O_8 都出现。

从图 10.11 可见，处理（X）在 O_4和 O_5 之间，而四种情况均有所增长，但是只有 A 和 B 的增长是处理起作用的结果。C 的增长是成熟的结果；而从历史上看，D 也不可能是处理的结果，因为 O_2 和 O_3 之间也有过增长。

10.6.4.2. 等值时间样本设计

和时间次序设计一样，等值时间样本设计（Equivalent Time-Samples Design）也适用于只有一个组，没有控制组。所不同的是，这个组接受经验的型式是事先决定的。这就是说，研究者必须让被试系统地接受处理。这种设计图式如下：

$$X_1 \quad O_1 \quad X_0 \quad O_2 \quad X_1 \quad O_3 \quad X_0 \quad O_4$$

这也是一种时间次序，所不同的是处理（X1）反复使用，而且交替使用另一种处理（X0）。这种设计满足了内部效度的要求，可以控制历史的影响，因此比时间次序设计优越。外部因素的影响不大可能在每次都和 X1 同时出现，所以我们只要把 O_1 和 O_3 的平均分与 O_2 和 O_4 的平均分加以比较，就可以排除历史的影响。这种设计还可以用来分析次序的效应：

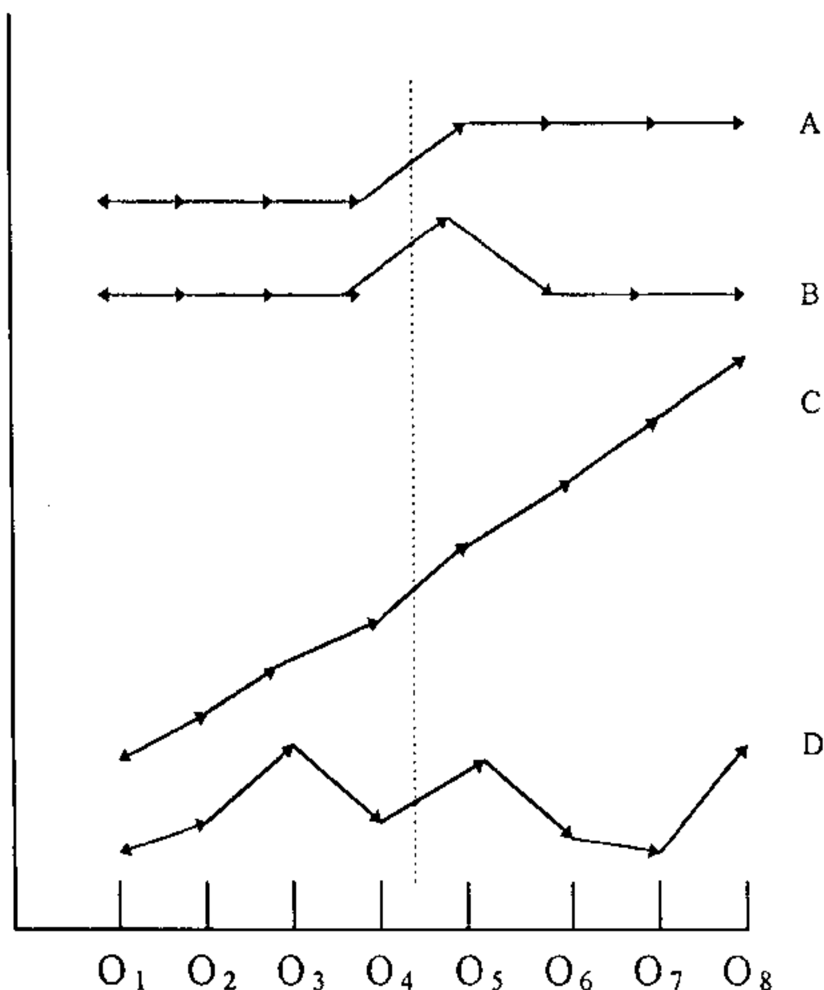


图 10.11 时间次序设计的控制

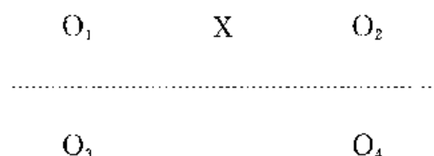
	第一次执行	第二次执行
X	O_1	O_3
X_0	O_2	O_1

例如一个英语教师想了解语言实验室的教学效果，但又没有办法组织控制组。他就首先让学生在正常的教室上课（X_1），然后组织了一次检测（O_1）；在第二周，他带学生到语言实验室上课（X_0），然后组织了另一次检测（O_2）；到第三周，他又让学生在正常的教室上课，然后再检测（O_3）；到第四周，又到语言实验室上课，再检测（O_4）。如果要了解这两种教学环境对教学的影响，我们可以把 O 1 和 O 3 与 O 2 和 O4 比较。如果要了解时间的影响，我们可以把 O_1 和 O_2 与 O_3 和 O_4 比较。在两组里用同一批被试可以排除选择的影响。

10.6.4.3. 非等值控制组设计

在教育实验中，我们往往不能随机安排被试。如果一个学校的校长同意拿出两个班来做实验，他不大可能同意我们把这两个班打乱，再重新组合两个班。非等值控制组设计

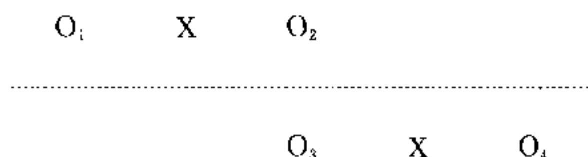
(Nonequivalent Control Group Design) 可以用来处理这种原封不动的非等值组：



除了不能做到随机安排被试这一点以外，这种设计和实验前后测试的控制组设计是一样的。为了保证非等值组在实验开始时是等值的，我们只好在实验前进行一次测试，比较它们依变量的分数（O_1 和 O_3）。由此可见，在这种设计里，实验前的测试比真正的实验还要重要，它是为了控制因为不能随机安排被试而产生的选择偏颇。

10.6.4.4. 分离样本的实验前后测试设计

在某些情况下，我们不能同时对被试都做实验处理，例如我们要培训 1000 人，但是设备条件只能容许每次培训 100 人。这个培训计划要延续下去，每次都换新的学员，然后重新再来。所有的学员都在某一段时间参加培训，我们就无法把他们安排为接受培训和不接受培训两种处理，进行比较。因为每一个人都只接受一次培训，我们也无法使用等值时间样本设计和时间次序设计。这就需要用到分离样本的实验前后测试设计（Separate - Sample Pretest-Posttest Design），其图式如下：



这实际上是把一组实验前后测试设计（O_1XO_2）重复做一次，其目的是控制历史影响。在一组实验前后测试设计里，可能有一些和 X 同时发生的事件会影响 O_2，使用分离样本的实验前后测试设计，我们就可以观察两次的结果，如果 $O_2>O_1$ 而 $O_4>O_3$，那么处理的效果就得到证明，因为外部因素不大可能会两次都和 X 同时发生。

但是这种设计可能会在三个方面失效：一是测试效应，因为它不能不使用实验前的测试；二是成熟影响，因为第二个组在开始时一般都会比第一个组成熟，如果培训时间为一年，第一个组的平均年龄为 18 岁，第二组就会是 19 岁；三是选择和成熟的交叉影响。

10.6.5. 事后的实验设计

事后（ex post facto）指的是实验者不做实验处理，而仅仅在这个处理自然发生时才去观察其效果。这时实验者把这种事后发生的效果看成是一个依变量。从图式上看，这种实验设计和上面所谈到的实验设计并无两样，所不同的是实验处理是通过选择，而不是通过操纵放进实验里的。因此我们不能简单地把自变量和依变量之间的关系看成是一种因果关系。

10.6.5.1. 相关研究

相关研究（Co-Relational Study）从一组被试中收集两套或更多套数据，以决定这些数据

之间的关系。这种方法可以表示如下：

$$O_1 \qquad O_2$$

例如我们认为外语学习的好坏和母语的水平有一定的关系，就可以使用相关系数的方法来观察它们的关系。如果相关系数比较高，就说明它们之间有密切相关，即母语水平高的人的外语也学得好，低的人则学得差。但是相关关系不一定表示两者的因果关系，可能有几种情况：

(1) O_1 所测量的变量引起 O_2；

(2) O_2 所测量的变量引起 O_1；

(3) 第三个没有测量到的变量引起 O_1 和 O_2。

由此可见，相关系数高不一定就是存在因果关系，但因果关系却意味着相关系数高。为了要了解因果关系，我们必须设计一个实验，在有实验控制的条件下观察一个变量。

10.6.5.2. 标准组设计

当我们在特定的环境中进行研究，并希望提出一些关于导致某些情况产生的假设时，我们往往需要把这些情况的特征和它相反的情况的特征加以比较，这就要使用到标准组设计(Criterion-Group Design) 的方法。

例如我们想了解有哪些因素可导致好的教学效果。在组织真正的实验之前，我们需要形成某些造成教学效果好坏原因的看法。在这个时候，我们使用标准组设计的方法，先采取一些评估方法来分出教学效果好和教学效果坏的两个教师组，然后再对比这两个组在课堂教学行为中的不同。研究者还可以观察这两个组的背景和教学能力，以了解有哪些因素可导致好的教学效果。标准组设计和相关研究是了解两个变量之间的关系的两种互为补充的方法。标准组设计可简单表示为：

$$\begin{matrix} C & \qquad O_1 \\ & \qquad O_2 \end{matrix}$$

我们使用 C，而不是 X，来代表按照一定的标准来选择的经验。

10.7. 观察与测量程序

10.7.1. 测量的信度

测量的信度指的是测试的一致性。一把橡胶做的尺子不可能是一把很可信的测量工具，因为橡胶有伸缩性，冬天测量的结果和夏天测量的结果会不一致。所以一个实验或测试的信度指它在重复测量时产生同样结果的程度。但是我们测量的对象是人，不是物，要想两次测量的结果完全一样，很难做到。因为人的因素很不稳定，测量中的误差很难避免。我们在一次

测量中所得的分数（X）并不完全是学生的真正的分数（t），换句话说，学生所得的分数中包含了一些误差（e）。这三者关系可以表示如下：

$$X=t+e \text{ 或 } t=X-e \tag{10.1}$$

这个公式表示的是对一个人的某种能力的测量，但是信度的概念指的是对不同人的反复测量，所以必须改写这个公式，使它能表示不同的人的X，t，和e。我们可用VAR（方差）来表示它们的差异，故有：

$$VAR(X)=VAR(t)+VAR(e) \tag{10.2}$$

如果我们在方程式的两边都除以VAR（X），便有：

$$1=\frac{Var(t)}{Var(X)}+\frac{Car(e)}{Var(X)} \tag{10.3}$$

因此信度（r）可以表示为VAR（t）和VAR（e）的比率：

$$r=\frac{Var(t)}{Var(X)} \quad \text{或} \quad 1-\frac{Var(e)}{Var(X)} \tag{10.4}$$

下图表示r和误差的关系：

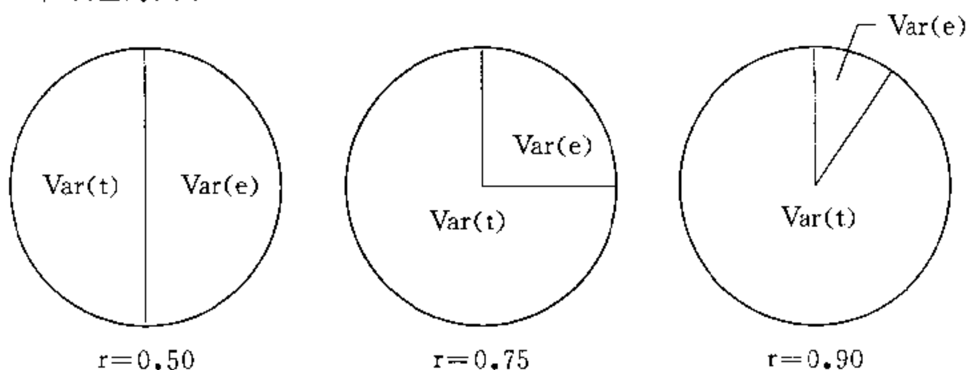


图 10.12 信度和误差的关系

从图中可看出，r值为0.50和0.75时，误差实在很大，故一般测量均要求信度在0.90以上。

由此可见，我们所说的测量信度指的是一个测试的分数的信度，而不是这个测试本身。而且由于我们把信度看成是不同的人的行为的函数，所以我们所说的信度就只能用来说明它所据以计算的那个样本。经典测试理论就是在信度概念的基础上发展起来的。

信度估算的方法有多种：

10.7.1.1. 再测法

这是估算信度的最简单的方法，用同一套试题在不同的时候测试同一批学生，然后再求出两次考试分数的相关系数。如果两次结果完全一样，其再测信度就是1.00。

但是这种方法也有它的问题：(1) 考生的水平很难说不会发生变化；(2) 考完一次试后，考生都会记住题目和答案，在第二次考试里，他们无非是重复第一次考试所做的东西；(3) 在实际的操作中，往往不易找到原来的考生。

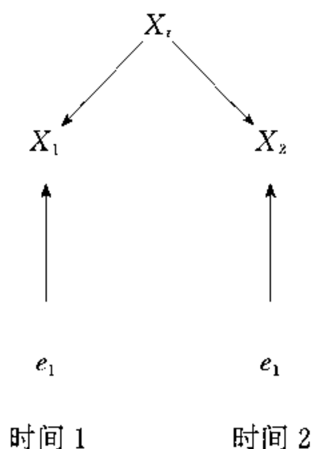


图 10.13 再测法的示意图

10.7.1.2. 平行试题法

这是以上方法的改善，用两套平行试题来代替同一套试题。平行试题应该在格式和内容上都保持一致。其他方面和上述方法相同。这种方法的好处是避免考生机械重复，但是考生的水平仍会发生变化，因此两次考试的时间不要相距太远。平行试题的主要问题是很难出两套真正平行的试题。这种方法的示意图[①] 是：

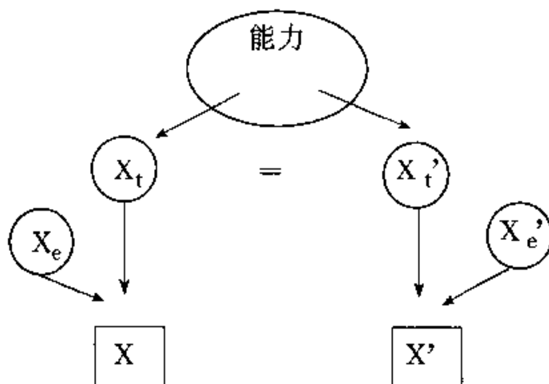


图 10.14 平等试题示意图

不管是再测法还是平行试题法，我们都是从真正的分数方差的角度来考虑信度，但是我们无法得知一个人的真正的分数，因此也无法得知信度，而只能从观察到的分数去估计。所以就把相关系数作为估计的基础。

但是在任何一种测试条件里，测量误差的来源是不只一个的。我们让几个组的学生去做听力测验，他们听到的是一些短文或对话的朗读或录音，然后从四个选择中选出正确的答案。我们假定学生的分数的不同反映了他们听力水平的高低。但是学生分数的差异也可能反映了他们参加考试当天的心情，或是他们对朗读或录音的感觉（如音量、速度、清晰度等等），或

① 这幅图是 Bachman（1990）所提出的。

他们对不同语篇的内容的看法。因此我们也可以从另一个误差源的角度来考虑计算考试分数的信度。一般来说，可以有三种方法：

(1) 内部一致性 (internal consistency) 估算。它主要是考虑测试内部和评分程序方面的误差源。

(2) 稳定性 (stability) 估算。它主要是考虑不同时期的测试分数是否一致。

(3) 等值 (equivalence) 估算。它主要是考虑一个考试的不同的试题所得到的分数如何等值。

后两个问题是大规模考试要考虑的专门性问题，我们不在这里展开。下面介绍的是内部一致性估算的几种方法。

10.7.1.3. 对半信度估算法

对半 (split-half) 法把试题分为两半，然后决定这两半的分数在多大程度上是一致的。我们实际上是把每一半试题看成是一个平行试题，因此必须有一些假定：这两半是等值的，它们的平均分和标准差是相等的；而且它们是相互独立的，即一个人在这一半的行为并不决定他在另一半的行为。这些假定并不意味着这两半没有关系，而只是说它们之所以有关系是因为它们测量同样的倾向或能力，而不是一半的行为决定另一半的行为。

最简单的对半法是把试题分为上下两半，但是语言测试是一种能力测试，往往把容易的题目放在前面，然后逐步加难。这样对分就使两半不等值了。另一种办法是把题目按单双数分为两半。但是在很多情况下，这种对分法并不很科学，因为我们不能用随机办法来安排题目。在一套多项选择题，不同的项目是用来测量不同的能力的：把一套试题随机地分成两半，其内容是不等的。应该按照所设计的内容来分成两半，以保证它们在内容上相等。还有一个问题是试题中的项目并非互相独立的，语言测试中的阅读题和完形填充题，都有一些并非互相独立的项目。

如果我们把试题分为两半，我们就可以使用 Spearman-Brown 的预测公式来计算对半信度系数：

$$r_{tt}=\frac{2r_{hh}}{1+2r_{hh}}$$

r_{tt}=试题信度系数 (10.5)

r_{hh}=两半的相关系

使用这个公式假定 (1) 这两半是平行的试题，其平均分和标准差是相等的；(2) 这两半在实验上是互相独立的。后一点很难验证，因为我们要说明考生在一半的行为并不影响另一半的行为。为了检验这个公式的有效性，最简单的方法是把试题做各种对半处理，例如我们有四个项目，我们可以把 (1) 1 和 2 作一半，3 和 4 作一半；(2) 1 和 3 作一半，2 和 4 作一半；(3) 1 和 4 作一半，2 和 3 作一半。然后各求出其信度系数，再取其平均数。但是这种方法很不实际，因为项目越多，各种搭配就越多。一个只有 30 个项目的试题竟有 77 558 760 种搭配[①]？

① 计算这种搭配的方法是 $C_n^{n/2}/2$，n 为项目数。

10.7.1.4. Kuder-Richardson 信度系数

好在有一种建筑在项目统计特征基础上的方法可以计算各种搭配的平均分。Kuder & Richardson 所发展的方法使用了试题的项目的平均分和标准差。一个两分法的项目（即不是对就是错）的平均分就是使用考生答对该项目的比例，表示为 p，而所有考生答错该项目的比例则为 1－p，或 q。而两分法的项目的方差则可表示为 pq。Kuder-Richardon 的公式（KR-20）建筑在项目方差的和与整份试题分数的方差的比率基础上：

$$r_{tt}=\left(\frac{n}{n-1}\right)\left(1-\frac{\sum pq}{s^2t}\right) \tag{10.6}$$

r_{tt}为试题的项目数，s^2t 为整份试题分数的方差，p 为一个项目的答对率，q 为它的答错率。Kuder-Richardson 还有一个简化的公式（KR-21），使用起来更方便，但不如前者精确。

$$r_{tt}=\left(\frac{N}{N-1}\right)\left(1-\frac{\overline{X}\ (N-\overline{X})}{Ns^2t}\right) \tag{10.7}$$

我们不难从这个公式中看出，信度的高低取决与试题的项目数的大小和标准差的大小。所以它只能用来计算客观性题目（如选择题和正误题）的信度。由此可见信度理论和客观性题目是一种唇齿相关的关系。

10.7.1.5. α 系数

上面的几种公式适宜于估算客观题的信度系数，但是一个试题往往还包括一些不是两分法的主观题，难以使用这些公式。Cronbach 提出的 α 系数就是为了解决这个问题的。在他的公式里，$\sum s^2i$ 就是一份试题的各大题的方差之和，而 S_x^2则是整份试题分数的方差。

$$\alpha=\frac{n}{n-1}\left(1-\frac{\sum s^2i}{s^2x}\right) \tag{10.8}$$

如果我们已经计算出各个大题的相关矩阵，我们可以先算出相关系数的平均数，然后然后再使用下面的公式：

$$\alpha=N\overline{p}/\ (1+\overline{p}\ (N-1)) \tag{10.9}$$

N 为大题数

$\overline{p}$ 为相关系数的平均

10.7.1.6. 阅卷员信度

一些主观性题目（如作文、口试）需要阅卷员来评阅，评出来的分数，常会有误差。一个阅卷员自己可能会不一致，这就是阅卷员内部信度（intra-rater reliability）的问题；几个阅卷之间也会不一致，这就是阅卷员之间的信度（inter-rater reliability）的问题。

1. 当一个阅卷员评阅主观性题目时，他必须使用一定的评分标准。如果他能保持这个标准的一致性，他就能评出一组可信的分数。当然这有一个前提，他所评阅的语言样本是没有误差的，所以阅卷员应用标准所得的结果就是学生的真正的分数。但是如果评阅的答卷数很大，就会产生一些前后次序的效应。例如在评阅作文时，阅卷员起初定的标准是注意内容和谋篇布局，可是他评阅时却发现越来越多的语法错误，使他无意识地注意语法错误，因而影

响他的评分。这样评分标准就无形中改变了。又如我们在评阅时，看到都是写得较差的答卷，当一篇较好的答卷出现时，我们就容易评出较高的分数。如果阅卷的时间较长，也会出现前紧后松的情况。

要了解一个阅卷员的评分信度，我们可以让他在不同的时候评阅同一份答卷，看评出来的分数是否一致。一般是随机地抽出一些试卷进行重评，然后计算两次评分的相关系数。另一种办法是计算 α 系数，把每一次评分看成是一个大题。我们可把每个人的两次分数组合在一起，然后分别求出每次分数的方差（s_{r1}^2和 s_{r2}^2）和组合分数的方差（s_{r1+r2}^2）。α 系数的计算公式如下：

$$\alpha=\frac{n}{n-1}\ (1-\frac{s_{r1}^2+s_{r2}^2}{S_{r1+r2}^2}) \qquad (10.10)$$

2. 阅卷员之间的信度。不同的阅卷员之间掌握的评分标准更不容易取得一致。要了解他们之间的信度，一般是让两个阅卷员评阅同一批答卷，然后计算其相关系数。如果阅卷员不止两个，就必须采用 α 系数计算方法：把每一个阅卷员看成是一个大题。计算其方差，然后计算出不同阅卷员的评阅分数的方差之和。（见 10.7.1.5）还有一种办法是把答卷随机分配给阅卷员，然后把每个人的评阅分数的分布和整体评阅分数的分布加以比较，并根据差异进行调整。（详细做法可参阅桂诗春，1986）

10.7.1.7. 标准误

根据所得的信度，我们就可以使用下列公式求出标准误（SE_m）：

$$SE_m=S\sqrt{1-r_{tt}} \qquad (10.11)$$

S 为标准差

r_{tt}为信度系数

例如标准差为 15，信度系数 0.9，我们就可以根据这个公式推算出一个标准误为 4.74。标准误可以用来表示一个考生的分数偏离他的真正分数的界限。如果一个人的分数为 50 分，那么在 95％的情况下，他的真正分数在 50±2 个标准误之间，即 59.48 到 40.52 之间。因为按正态分布的原理，在标准分 0.96 之间的面积刚好是 95％，为了方便，我们取 2。

10.7.1.8. 影响信度的因素

影响信度的因素有好几个方面：

1. 试题的长度。试题长，项目多，信度就会提高。
2. 试题的同质性。如果一份试题考的都是同一能力倾向，信度也会提高。
3. 试题的区分度。区分能力强的项目越多，信度也会越高。
4. 考生的差异性。如果考生能力差异很大（表现为标准差很大），信度也越高。
5. 有足够的考试时间。目前所使用的信度公式，是以考生能够做完所有的题目为前提的，因此它用于估量能力考试的精确性要比估量速度考试的高。用对半法来估量一个速度考试会很不准确。

10.7.1.9. 经典测试理论估算信度的问题

经典测试理论的真正分数模型有不少令人不满意的地方。主要是测试的误差来源很多，而且交互起作用，经典测试理论把错误的方差的来源看成是同质的。它假定测量误差对所有的考生都是一样的，因此只需要用一个量度就能描述考试的信度。另外，作为估算信度的基础的是平行试题，但实际上两份试题很难做到真正的平行。

针对经典测试理论的问题出现了两种测试理论：概化理论（Generalizability Theory）和项目反应理论（Item Response Theory）。概化理论是由 Cronbach 等人提出的，这是一个观察不同来源的方差的相对效应的模型。这个模型以因子设计（factorial design）的框架和方差分析为基础。概化理论把某一个分数看成是一个假设总体的各种可能的分数的一个样本，当我们在解释一个分数时，我们是在把一个量度概化为全部量度。换句话说，我们可以把一个人在一个测试中的行为概化为他在别的场合里的行为。这个样本的行为（表现为分数）越可信，它的概括程度就越高。所以信度问题就是概化问题，我们对一个分数的概括程度取决于我们是怎样定义全体的分数。概化理论的应用分两步：第一步是概化研究（G-study），根据分数的用途设计一个观察我们感兴趣的方差来源的研究，包括决定有关的方差来源（如倾向、测试方法、个人属性、随机因素等等），设计一个方案去收集那些足以使我们区分方差源的数据，根据这个方案来组织考试，然后进行合适的分析。在这个概化研究的基础上，我们就可以估计不同的方差源的相对大小（称为"方差成分"）。第二步是决策研究（Decision Study），在决策研究中，我们在可操作的条件（即使用测试来作出决策）下使用概化理论的程序去估计方差成分的大小。

但是经典测试理论也好，概化理论也好，都不能预测某一个人做某一个项目的情况。这是因为（1）它们并不假定一个人的能力水平会影响他在测试中的行为；（2）唯一能够预测一个人在一个测试中的行为的是难度指数 p，而这仅是项目的答对率。所以唯一能够预测一个人答对一个项目的是该项目的平均分。针对这些缺陷，心理测量学家又提出了项目反应理论（贡献最大的是 Lord，1980）。项目反应理论认为，一个人答对一个项目的期望行为是项目的难度水平和他本人的能力水平两者的函数。项目反应理论有一些假设（如单维度假设、局部独立性假设，等等），这里不予展开。单就信度而言，它提出信息函数的概念，项目的信息函数指的是一个项目所提供的估计一个人的能力水平的信息，测试的信息函数指的是所有项目信息函数之和。而测量误差是根据信息函数来估算的，这样一来，每一个项目都可以有自己的测量误差，每一个能力水平也可以有自己的测量误差。因此项目反应理论所提供的信息就比其他理论所提供的要多些。

10.7.2. 测量的效度

用简单的话来说，效度就是一个工具测量它所要测量的东西的程度。但是我们所要证明的不是测量工具本身是否有效，而是测量工具用来测量某种东西是否有效。一把用来秤米的工具可以是有效的，但是用来它秤金子就不一定那么有效。所以效度实际上指的是我们所获得的证据在多大程度上支持我们根据分数所作出的推断。有效与否指的是我们使用一个测试来所作的推断，而不是这个测试本身。

信度和效度是互为补充的。考察信度是为了回答这样的问题："考试分数中有多少方差是测量误差引起的？"和"有多少方差是测量误差以外的因素引起的？"测量误差以外的因素所引起的误差也可以叫做"可信方差"。考察效度是为了回答这样的问题："有哪些能力可以说明考试分数中的可信方差？"所以我们可以说，信度考虑的是考试分数中有多少方差是可信方差，而效度考虑的是哪些能力会导致可信方差。Campbell & Fiske (1959) 指出："信度是两种使用尽可能相似方法去测量同一种倾向的企图的一致性。效度是两种使用尽可能不同方法去测量同一倾向的企图的一致性。"这个差别可以表示为下图：



图 10.15 信度和效度测量同一倾向一致性的差别

10.7.2.1. 内容效度

测试的目的是为了了解考生在一些实际环境中是如何运作的。但是我们不可能在测试中再现所有的环境，而只能选择一些样本，然后根据考生在这些样本中的表现，评估其运作的能力。如果我们所测试的样本能够充分地代表总体，我们就可以说测试在内容上是有效的。例如我们出一道作文题《试论滑雪的好处》来考学生的写作能力。我们的考生遍布全国，有相当一部分南方的考生连雪都没有看见过，更没有滑雪的经验，要他们去讨论滑雪的好处当然就写不出很多东西。这样我们的作文就没有考出它所要考的写作能力，这是因为我们定的题目缺乏代表性。

内容效度 (content validity) 包括两个方面：

1. 内容的关联性。这牵涉到规定能力范围和测试这些能力的方法。我们需要回答这样几个问题：(1) 测试要测量的是什么？(2) 我们用来测量考生的项目有些什么属性？(3) 考生可能会作出什么回应？

2. 内容覆盖范围。测试要求考生所完成的任务在多大程度上反映能力范围的要求。在某个意义上说，如果我们能够有一个关于能力的细目表，我们就可以使用随机抽样的办法，选择一部分有代表性的项目。问题是这种细目表很难制订，即使是教学大纲和考试大纲都很难全部列出。要使测试项目有代表性，还有一个数量的关系，例如有一道考英语语法的大题，总共只有 10 道题，其中 7 道是考介词的。我们就很难说这个大题有代表性，即使是它考的范围较广，到底题量（样本）太少，难以反映语法的全部内容。所以效度和信度是密切联系在一起的。

内容关联性和内容的覆盖面应该综合加以考虑。例如一些如写作和问答题的主观性题目有较好的关联性，但是受了时间的制约，一份试题不可能包括很多写作题，结果覆盖面就会太窄。所以主观题和客观题要配合起来使用，以提高内容效度。

10.7.2.2. 预测效度

预测效度(predictive validity)和共时效度(concurrent validity)都属于与标准有关(criterion related)的效度。因为这种效度表示了考试分数和某些标志考生能力的标准的关系。预测效度表示的是某些标志考生将来的能力的标准。例如我们可用一个语言测试的分数来预测考生将来胜任某种工作或取得好的学习成绩的能力。我们可以先给学生一次作文考试、然后把学生分到不同水平的写作班级，再把第一次分班用的作文考试分数和学生后来的写作分数来比较（如计算相关系数），就可以了解第一次考试的预测效度。Thorndike (1977) 曾经把美国空军驾驶员的学能考试和经过一段时期学习后的考试不及格率作比较。学能成绩分为 9 等，结果级别越低，不及格率就越大，它们的相关系数为 0.49。

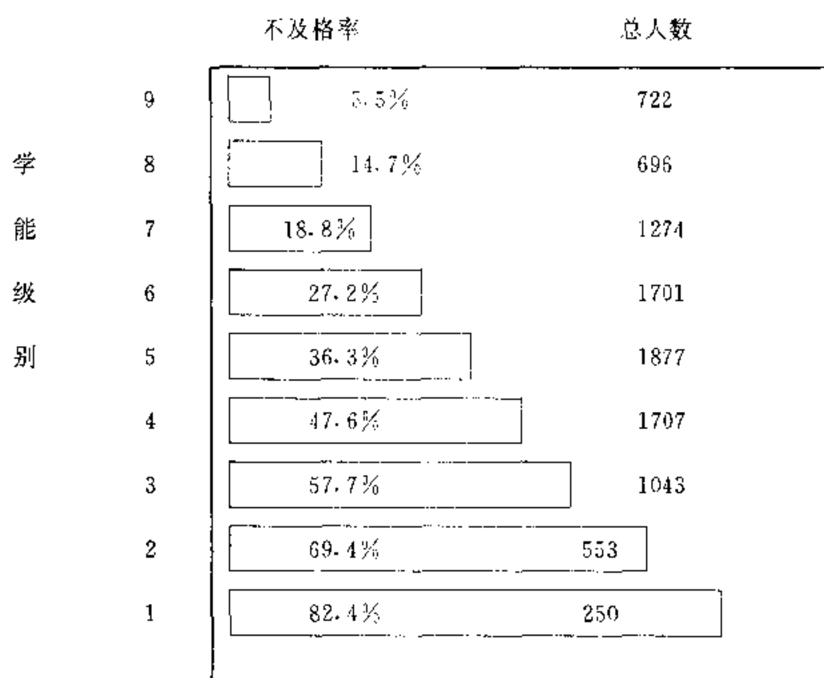


图 10.16 学能考试的预测效度

1986 年，我们也曾把 1985 年通过高校入学考试考入广州外国语学院就读的学生的英语科成绩和他们 1986 年的期末成绩相比较，笔试成绩的相关系数为 0.468，口试成绩的相关系数为 0.574。这就是英语科分数的预测效度。

10.7.2.3. 共时效度

共时效度表示的是一个考试分数和另一个同时使用的标准的分数的关系。这是在建立标准化考试的过程中经常使用的一种手段。例如我们在 80 年代建立供选拔出国人员使用的英语水平考试 (EPT)，为了了解这个考试的标准化的程度，我们让考生差不多在相同的时候参加了两次 EPT 和两次 TOEFL 考试，然后了解它们的共时效度。

TOEFL 是一个标准化的考试，所以经过共时效度的验证，我们也可以说当时的 EPT 也具有标准化的特征。

表 10.11　　**EPT 与 TOEFL 的相关矩阵**

	I. EPT 样题	II. EPT (M1)	III. TOEFL 样题	IV. TOEFL (80/12)
I.	—	0.86	0.85	0.85
II.	0.86	—	0.86	0.86
III.	0.85	0.86	·	0.87
IV.	0.85	0.86	0.87	—

10.7.2.4. 构想效度

构想效度（construct validity）要考察的是一个考试的结果在多大程度上和我们根据某一理论所作出的预测相一致。它验证的是我们所作的假设是否有效。其实所有的测试都牵涉把测试分数解释为能力的标志，一当我们提出这样的问题：这个测试究竟测量了什么？就需要考察构想效度。所以构想效度是一个把其他的效度统一起来的概念。在心理语言学里所作的各种实验都有一个构想效度的问题。

在验证构想效度时，我们经常使用的是相关方法，并在这个基础上进一步使用诸如因子分析等多维度分析方法①。但是相关方法有时会引起一些含混。例如我们发现关于连贯性的一个多项选择题和另一个关于结构组织的多项选择题有较高的相关，这有三种可能：(1) 它们的分数都受到一个共同的倾向（语篇能力）的影响，如图 10.17 的 (a)；(2) 它们的分数都受

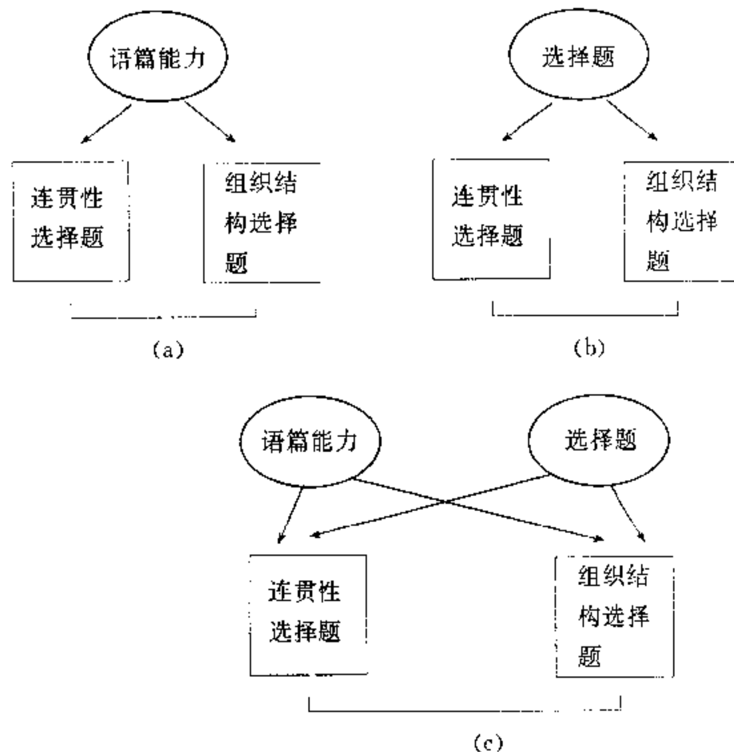


图 10.17　测试方法和所测试的倾向的几种关系

① 因子分析方法可参看 11.7.3.；多维度量表方法可参看 11.7.5.。

到一种共同的测试方法（多项选择题）的影响，如图 10.17 的（b）；（3）它们的分数既受到倾向，也受到方法的影响，如图 10.17 的（c）。

Campbell & Fiske（1979）因此提出一个多维度多方法矩阵设计（Mulitdimensional-multi-method Matrix Design，简称 MTMM）来解决这个问题。

在这种设计里，每一个量度都是一个倾向和方法的组合，它还包括了把多倾向和多方法组合在一起的测试。它的好处是可以考察各个相关的聚合（convergence）和区别（discriminaion）的型式。聚合指的是上面谈到的共时的标准，考察的是使用不同方法测量同一倾向的一致性程度。区别指的是使用不同的或相同的方法测量不同的倾向产生不同结果的程度。聚合可以表现为不同方法测量同一倾向的高相关；而区别可以表现为不同方法测量不同倾向的低相关或零相关。如果用同一方法测量不同倾向的相关很低，也算是区别。MTMM 的设计可以表示为图 10.18。图中表示的是用两种方法（选择题和写作题）测试两种倾向（语篇能力和社会语言能力）。这种设计可用结构方程模型[①] 来进行分析。

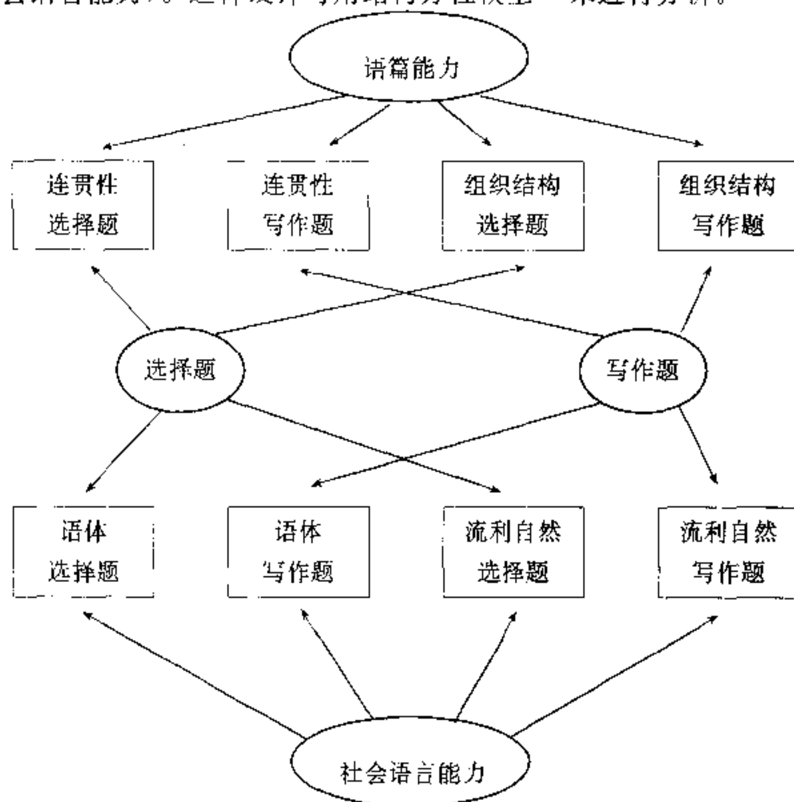


图 10.18 MTMM 示意图

Bachman 指出，从 MTMM 所得到的数据可以用不同的方法处理，（1）直接观察聚合和区别的相关系数；（2）使用方差分析的方法；（3）验证性因子分析。

① Structural Equation Modeling，可参看 12.7.7.2。

10.7.3. 数据量表

所有的量化数据都是可以观察的，问题是它们的观察和测量的方式不一样。英语水平这个变量可以通过一个考试来测量，并表示为一组分数。而诸如性别和民族这样的变量却需要学生自己提供信息，然后由调查人把信息归到不同的范畴，每一个范畴里有多少人。前者是一组分数，而后者却是一套范畴。

在统计中，我们使用不同的量表来记录这些测量方式不同的数据。量表无非是观察、组织、分配数据的不同方式的名称。从精确度而言，不同的量表对待数据有不同的要求。

10.7.3.1. 量表类型

1. 名义量表（Nominal Scales），对变量中的数据进行分类，并给予名称。数据的归属可以是自然的，如性别、民族、母语、喜爱态度等等；也可以是人工的，如在一个实验里把学生分到控制组和实验组。一个名义量表有两个或更多的范畴，数据不是属于这个，就是属于那个范畴。名义量表的数据就叫做名义数据，也可称为离散性数据（discrete data）、范畴性数据（categorical data）或非连续性数据（discontinuous data）。这些范畴中的数量没有数学的联系。一般我们采用条形图（bar charts）来表示名义量表，如我们调查了两个年级的男女生的数目，可表示为：

2. 顺序量表(Ordinal Scales)，把数据排列成为次序，例如我们可根据期末考试成绩把一个 30 人的班级排成高低次序，最高分的是 1，其次是 2，最后一名是 30。顺序量表使用了多于和少于的概念，它比名义量表提供的信息要多一些。当范畴只有两个时，名义量表和顺序量表的界线就划不清了，例如我们可以把学生的智商分为高低两组，这可以是名义量表；但如果把高低看成是一个次序，那就是顺序量表。在统计上，一般我们把只有两个范畴的量表看成是名义量表。

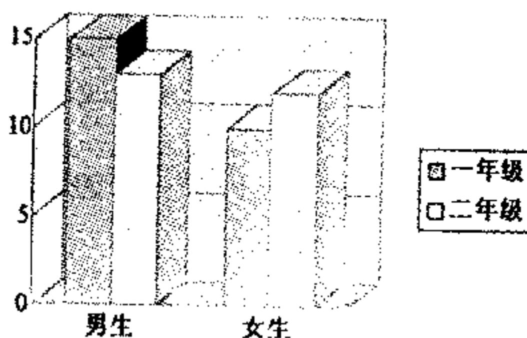


图 10.19 两个年级的男女生数目条形图

3. 区间量表（Interval Scales），使用得最广的量表。它不但告诉我们事物的次序，而且还提供事物之间的距离。假定在一次课堂考试里，考第一名的得 95 分，第二名的得 85 分，第三名的得 75 分。他们的距离都是 10 分，这就是说，第一名高于第二名的程度相当于第二名高于第三名的程度。如果在另一次课堂考试里，考第一名的为 95 分，而第二名的为 93 分，第三名的为 75 分，情况就很不一样。尽管从顺序量表上看都是一到三，但是第二次考试里的第一名和第二名很接近，他们却比第三名好很多。在区间量表里，每一分之间都是等距的，所以在应该 100 分的量表里，12 和 14 的距离为 2 分，而 98 和 100 的距离也是 2 分。评分量表和测试使用的是区间量表，而且还可以把原始分数转换成标准分，以保证区间量表的特性。区间量表一般表示为直方图（histogram）。

4. 比率量表（Ratio Scales），在物理中应用得较广，在行为科学中则很少采用。比率量

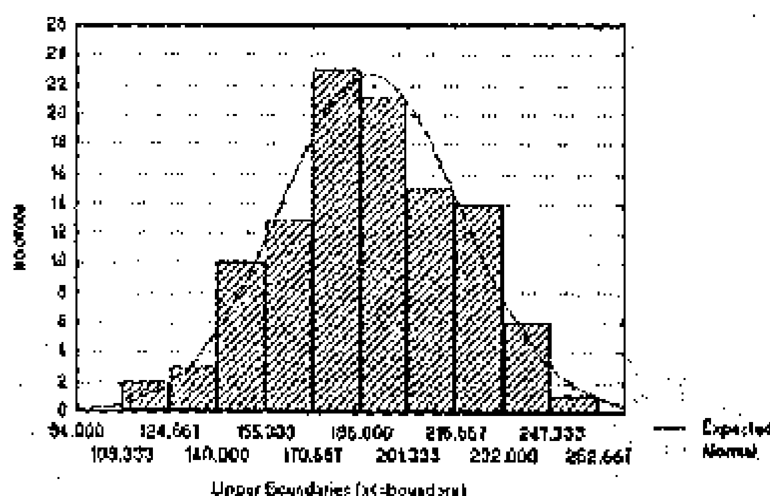


图 10.20 一个频数表的直方图

没包括一个真正的零值，也就是说，在量表上有一点是表示完全缺乏所测量的特征的。比率在量表上的各个位置是可以比较的，例如9欧姆是3欧姆阻抗的3倍，而6欧姆则是2欧姆阻抗的3倍，因此它们是相等的。但是区间量表则不同，例如在智商量表上120分的人和100分的人（相距为20分）比144分的人和120分的人（相距为24分）更为接近。但是从比率上看，它们是一样的（100/120 = 120/144）。这是因为智商量表是区间量表，而区间量表是没有真正的零点的。一般来说，在行为科学里，很少用到比率量表，所以以后我们就不再提到它了。表10-12是几种量表的一个简单的归纳：

表 10.12　　四种量表的比较

量表	所提供的信息内容			
	范畴（名称）	顺序（排列）	区间（距离）	零点，可相乘（小倍比）
名义	×			
顺序	×	×		
区间	×	×	×	
比率	×	×	×	×

5. 量表的转换。如果一个研究者想测量一组学生的学习兴趣，他可以采取不同的方式来收集数据：

（1）他可以把数据归属为不同的范畴：有兴趣和没有兴趣。

（2）他也可以按照兴趣的强度来排列次序。

（3）他也可以在一个兴趣的量表上给每一个学生评分。

假定他决定使用第三种方式收集数据，即区间数据，他很容易把区间数据转换成顺序数据，或把学生分为两半，即为有兴趣和没有兴趣两组，成了名义数据。由此可见，量表的转换是单向的，我们可以从高到低地转换数据，但反过来就做不到。例如我们按照平均绩点把

学生分为高、中、低三档，我们就不能再把这个名义量表转换成顺序量表，因为这三个范畴里并没有学生次序的信息。虽然我们可以从高到低转换数据，但我们必须认识到两点：一是在转换过程中，一些有用的信息会丢失；二是我们对基本变量性质的假设也可能会改变。例如我们把学生在期末考试里的分数从区间数据改为顺序数据，我们保留学生成绩的次序，但是学生分数之间的距离的信息丢失了。而我们对待期末考试这个变量的看法也改变了：我们只管学生成绩的次序，不管学生之间的距离的大小。

10.7.3.2. 量表的建立

量表是研究者用来量化一个被试对一个变量所作的反应的的装置。量表可以用作收集任何被试或任何对象的态度、判断或感觉的区间数据，这一般叫做态度量表。这种量表主要有几种：

1. Thurstone 量表。这是 Thurstone 等人早在 20 年代提出的，他们首先准备了约 20 个关于战争、死刑、教会、爱国主义、审查制度等等态度量表，收集了数百条态度语。第二步是请一些熟悉这些题目的人担任评判员，按照他们赞成的程度把这些态度语归为几个组（一般是 11 组）。这些评判员无需表明他们的态度，只要把态度语归类即可。第三步是把每一态度语作出次数分布表，取其中位数作为该态度语的量表值。第四步选择评判员的评分信度较高、且 0—11 之内都有相应量表值的态度语数十句构成一个态度量表。第五步是组织测验，让被试对量表中各态度语表示赞成或反对。第六步是评分，将被试表示赞成的的项目依量表值高低排列，求出中位数，即为被试的分数。

2. Guttman 量表。这是 Guttman 于 40 年代提出的，这个方法亦称为量表图分析法(Scalogram Analysis)。这种量表把被试各个时候的数据组合在一起，以观察其变化。例如我们可以按时间把一个儿童这 5 周内习得 5 个语素的次序排列如下：

表 10.13　　一个儿童 5 周内习得 5 个语素的排列

星期	语素				
	M5	M4	M3	M2	M1
5	0	1	1	1	1
4	0	0	1	1	1
3	0	0	0	1	1
2	0	0	0	0	1
1	0	0	0	0	0

从这个表可看出，在第一周他一个语素都没有掌握，在第二周他掌握了 M1，在第三周，他又掌握了 M2，……到第五周，只剩 M5 没有掌握。我们也可以在同一个时候，调查 5 个儿童，表 10.14 表示 S1 掌握了 M1 到 M5，是能力最强的被试；S2 掌握了 M1 到 M4，次之，等等。如果我们按列来看，则 M1 最容易，而 M5 最难。这是一个理想的量表，根据这个量表我们可以预测儿童习得语言的行为，如果 S3 掌握 M3，我们可以预测他也掌握了 M1 和 M2。

表 10.14　　**5个儿童掌握5个语素的排列**

被试	语素				
	M5	M4	M3	M2	M1
S1	1	1	1	1	1
S2	0	1	1	1	1
S3	0	0	1	1	1
S4	0	0	0	1	1
S5	0	0	0	0	1

但是并不是每一种数据都符合这种量表的。就项目而言，它们在难度上存在一定的关系，M5是最难的，它比M4难，而M4又比M3难。就被试而言，他们在水平是亦存在一定的关系，S1比S2的水平好，而S2又比S3好。如果所有的项目是同等难度的，而所有的被试有是同一水平的，我们就不可能有这个量表。例如所有的被试都是初学者，而所有的项目都很难，量表就会全部是零；如果所有的被试都是水平高，而所有的项目都很易，量表就会全部是1。实际的情况是往往像下表那样：

表 10.15　　**5个被试回答5个问题的实际情况**

被试	问题					
	难					易
	Q6	Q5	Q4	Q3	Q2	Q1
S1	1	(0)	1	1	1	1
S2	0	1	1	(0)	1	1
S3	0	0	(0)	1	1	1
S4	0	0	0	1	(0)	1
S5	0	0	0	0	1	1

在这个表里，就4个地方偏离了正常的情况，本来S1是水平最好的，是唯一掌握Q6的被试，因此我们预测他应掌握Q5，但在再产生他的行为时，却出现了错误。这可以说是再产生性错误（error of reproducibility）。对一个量表的全部错误的量度就是再产生性指数（C_{rep}），其计算的公式是：

$$C_{rep}=1-\frac{\text{错误总数}}{\text{反应总数}} \qquad (10.12)$$

即 1－4/30＝0.87。一般来说，再产生性指数应该在0.90以上，如果低于0.90，就意味着缺乏预测性。这仅是Guttman量表的一个简单的介绍，读者如对它在语言教学中的应用感兴趣，可参看Hatch & Farhady（1982）。

3. Liskert 量表。Thurstone量表的建立比较繁杂，而Guttman量表的应用面又不那么广，所以比较多的态度量表都使用了Liskert量表。Liskert量表是一个5度量表，每一度的区间都

是等距的。这个量表用来记录被试对一个态度语的赞成或反对的程度。例如：

学校教育应该强调素质的教育

很同意	同意	未决定	不同意	很不同意

被调查人在量表的相应的字眼上面做一个记号，以表示自己的选择。下面是对美国综合高等中学进行职业教育的一个调查表的几个项目：

综合高等中学的职业教育

这个量表供您表示自己对综合高等中学的职业教育的意见。请在左边的字母上面划一个圈，以表示您对每一态度语的意见。(SA 为很同意，A 为同意，U 为未能决定，D 为不同意，SD 为很不同意)

SA A U D SD 1、综合高等中学应该包括职业教育。
SA A U D SD 2、综合高等中学的受职业教育的学生有机会参加各种活动。
SA A U D SD 3、接受职业教育的学生不会觉得他们能适应综合高等中学。
SA A U D SD 5、职业课程不能避免综合高等中学中的学生淘汰率。
SA A U D SD 11、综合高等中学的职业教育课程的质量不会很高
SA A U D SD 17、每个中学毕业生都应该接受职业教育。

读者不但发现有些项目是从正面提问题的（如 1、2、17），有些是从反面提问题的（如 3、5、11）。因此打分的办法应该不同：从正面提问题的，按下列方法记分：

SA＝5，A＝4，U＝3，D＝2，SD＝1

从反面提问题的，则是

SA＝1，A＝2，U＝3，D＝4，SD＝5

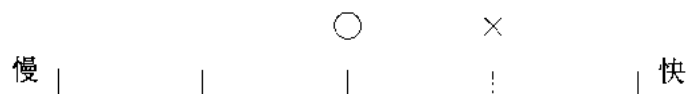
这样总分才能反映对所提问题所持的肯定态度。一个赞成在综合高等中学中进行职业教育的人会对正面的问题持赞成态度，对反面的问题持不赞成的态度；而一个不赞成在综合高等中学进行职业教育的人则会对正面的问题持不赞成态度，对反面的问题持赞成的态度。

所有的设计好项目都必须先拿去一个试点组试用，然后把每一个项目的反应和整份问卷的反应的总分来做相关分析。这种分析能够使我们了解每一个项目反映整份问卷的程度，然后从中拿出相关最高的 20 个项目来组成一份问卷。这份问卷里的正面问题和反面问题的数量应该大致相同。

10.7.3.3. 语义微分分析

语义微分（Semantic differential）原来是 Osgood 等人（1957）所设计的语义实验，用来了解人们是怎样赋予意义的。一个概念有各种语义特征，就好象一个人群有各种各样的人一样，这些语义特征有共同的，也有不同的地方，例如“快”、“慢”、“子弹”、“跑步”、“懒散”、“火箭”这些概念都很不相同，但是都可以统一在一个维度上：速度。如果要画一条线，

那么可以从零点算起，有不同的距离。这些概念在这根线上的位置表示了它们的相对的“意思”，例如“火箭”（×）和“子弹”（○）的位置可能是这样表示的。



Osgood 等人在实验中把一维的概念扩展到多维，例如他让被试对“毒药”（poison）和“笑”（laughter）两个词在下列量表中标出他们的判断。下面是一份典型的答卷：

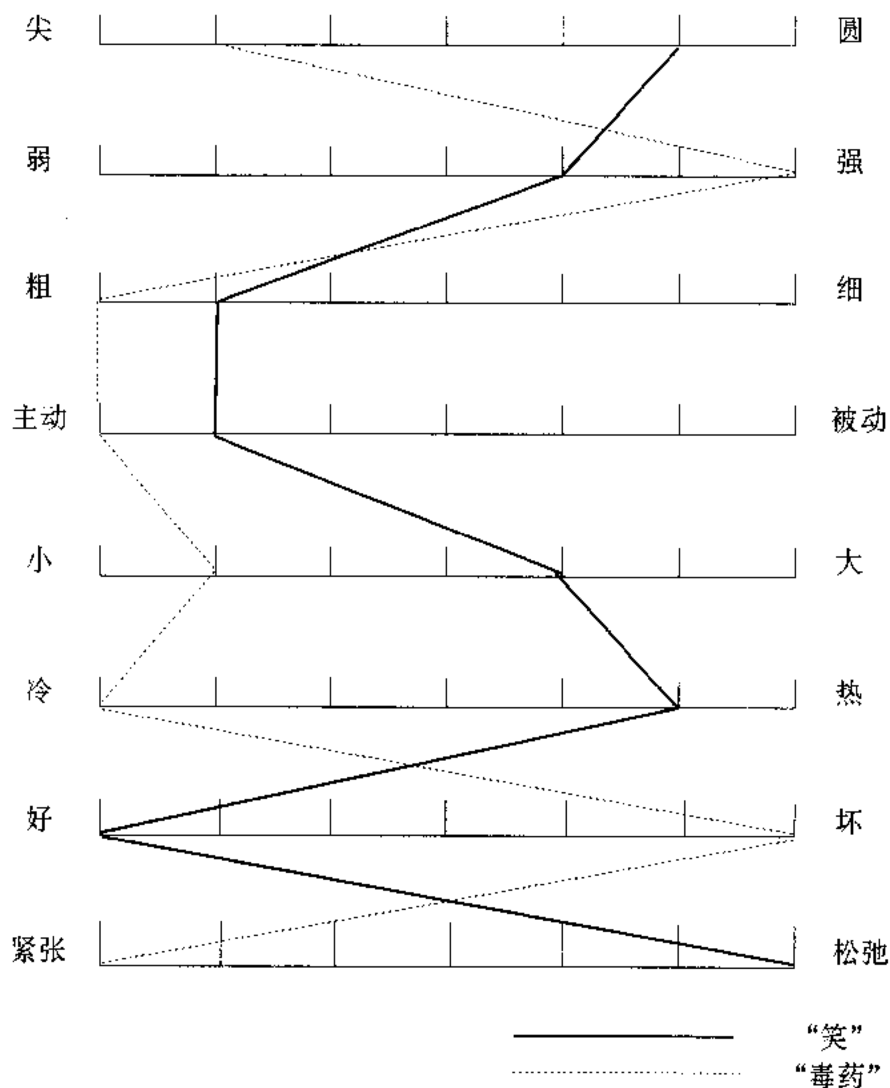


图 10.21 语义微分量表

在每个被试作出评判后，试验者把数据收集起来进行因子分析（factor analysis），可以归纳出 3 个共同的因素：

评价（evaluative），如：好—坏、美—丑、清洁—肮脏

强度（potency），如：强—弱、硬—软、大—小

活动（activity），如：快—慢、冷—热、紧张—松弛

实际上Osgood等人所测量的是人们对一些词的反应，而不是词的意义本身。Tuckman把这种方法引进到教育测量，他设计了一个关于“教师”的调查表，使用的也是这样的方法，调查表共有28个形容词的对子，如独创性—传统、耐心—不耐心、冷淡—热情、敌对—友善、创造性—常规性、公平—不公平、没有组织—有组织，等等。这些形容词对子实际上分为(1) 创造性；(2) 动力；(3) 组织性；(4) 热情与接受性4个维度。

10.7.4. 记录观察结果的手段

记录观察结果的手段有两种：一种是评分量表；第二种是编码系统。

10.7.4.1. 评分量表

评分量表是观察者归纳他对所观察的活动或行为所作的判断的一种工具。这种量表可以是3，5，7，9，100乃至无限个单位（最常用的是3，5，7个单位)。在量表的两头都有一些形容词的陈述语。观察者使用这些量表来记录一段时期的观察，使结果得以量化。下面是几个例子：

1. (5) (4) (3) (2) (1)

_____：__×__：_____：_____：_____：

大部分时间是教师在说话 　　　　大部分时间是学生在说话

2. (7) (6) (5) (4) (3) (2) (1)

_____：_____：__×__：_____：_____：_____：_____：

学生经常提问 　　　　学生很少提问

3. (9) (8) (7) (6) (5) (4) (3) (2) (1)

_____：_____：__×__：_____：_____：_____：_____：_____：_____：

教师作出所有的课堂决定 　　　　教师帮助学生作出课堂决定 　　　　在课堂上不作任何决定

4. |←——X毫米——→|

__________×__

教师经常维持课堂秩序 　　　　教师很少维持课堂秩序

5. 请用从0到100的整数来表示教师在课堂上说话的时间（100=所有时间，0=没有)。

__30__。

当我们把人作为测量工具来完成评分量表时，他们的观察往往会受到各种各样的影响。其中一种影响叫做“光环效应”（halo effect），就是说，如果观察者喜欢一个人，他容易在所有的量表上都作出肯定的反应。在样一来，量表测量的仅是观察者一般的肯定态度。如果在一些无大关联的量表上发现很密切的关系，我们就应怀疑评分会受光环效应的影响。由于评分的是人，都难免会出现一些不一致的地方或误差，这就会导致工具影响到内部信度。这时就要计算评分人之间的信度，通常的做法是让两个或更多的人独自完成量表，然后计算其相关系数，如果系数在 0.70 以上，评分员之间的误差就算是可以接受的了。更好的办法是把他们的评分相加，再取其平均数。

10.7.4.2. 编码系统

编码系统是另外一种记录事先决定的行为的手段，它把行为分为一系列的范畴，然后在这些行为出现时，就记录在相应的范畴里。和评分量表一样，编码系统也是为了把行为量化；所不同的是，评分量表是回顾式的，记录的是已经发生的、保留记忆中的东西；而编码系统是即时性的，一观察到有关行为就记录下来。评分量表用全面的方式归纳行为发生的频率；而编码系统则是按事先设计的编码来记录特定的行为的发生频率。

编码系统有两种。第一种是使用一系列的行为范畴，一旦这些行为出现，就马上记录在相应的范畴里面。什么时候编码取决于事先决定编码的行为，例如编码系统中有“建议”的编码范畴，一旦观察到建议的行为，就要把它记录下来。这种系统可以说是符号性编码系统。

另一种是时间性的编码系统。它也使用一系列的行为范畴，但是编码的不仅是事先决定的行为，或符号性行为，而是以时间为单位，把事先决定的一段时间内发生的所有的行为都记录下来。例如把每 3 秒种发生的所有的行为都记录下来。

和评分量表相比，编码系统有其自身的优缺点。优点是所收集的数据是“硬数据”，较接近现实生活。缺点是很难训练记录员，不容易取得很高的信度；而记录很花时间。所以除非有一个很好的编码系统，一般都使用评分量表，因为它较易操作。

10.7.5. 问卷与访问设计

研究者使用问卷和访问把所得到的关于人或事物的信息转化为数据。用这种方法可以测量人们的知识或信息、喜爱和厌恶，对事物的态度等等。它们也可用于了解发生过什么事情，现在正在发生什么事情。这些信息可以通过量表或评分量表转化为上面所谈到的数字和量化数据。

问卷和访问通过直接向人询问，而不是观察来获得数据。但是这种自我报告的方法也有本身的一些问题，因为（1）被调查人在完成问卷和被访问时必须合作；（2）他们必须如实报告自己的想法，而不是说一些他们认为应该如何的意见，或是说一些他们以为调查者喜欢听的意见；（3）他们必须知道自己有些什么感觉和想法，才能报告出来。

在准备问卷和访问时，研究者必须十分谨慎小心。他们必须经常应用如下的准则：

1. 一个问题在什么程度上会影响回答人尽量表现自己？
2. 一个问题在什么程度上会影响回答人通过预测研究者想听什么、想发现什么而进行不

必要的帮忙？

3. 一个问题在什么程度上会问一些回答人自己并不确切知道的信息，或是自己不大可能了解的信息？

问卷和访问的效度取决于上述三种考虑。但是有些信息是除了直接向被调查人了解外不可能获得的。有时，即使是可以用别的方法获得，也不如直接了解那么有效。在使用问卷和访问来获得数据时，它们的优点和问题都必须事先考虑清楚，然后决定取舍。

10.7.5.1. 应该怎样提问？

在访问调查里，有某些问题和回答的格式是经常采用的。我们先从提问方式谈起。

1. 直接的和间接的问题。直接的和间接的问题的区别在于问题在提取信息时的明显程度。如果我们问的是你是否喜欢自己的工作，这是一个直接的问题；如果我们问的是你觉得你的工作怎样，或是你觉得你的工作的某些方面怎样，然后再去根据回答进行推断，这就是间接问题。我们之使问题问得不那么明显，即使用间接问题，是希望能够提取一些较坦诚的回答，但是往往需要提较多的间接提问。要想得到一些坦诚的回答，也可采取别的办法，这在后面还会谈到。

2. 特定的和非特定的问题。一种问题和特定的对象、人或观念有关，我们所要提取的是对它们的态度、信念或想法。例如对一幅特定的画的态度。另一种问题则是探究一个较一般的领域，例如对图画的风格是否属抽象派的态度。一个访问者可以向一个学生了解他喜爱某一个教师的程度，或是他对一个班级（也就是这个教师在任教的班级）的满意程度。特定的问题也象直接提问一样，会令被调查者更为小心谨慎，因而给出不太坦诚的回答。非特定的问题往往要转弯抹角才能取得所需的信息，但不会吓到被调查者。

3. 事实和意见。所调查的问题可以要求被调查者提供事实，也可以要求提供意见。如果我们问学生使用什么教科书，这是就事实提问；如果我们问学生喜欢哪一种教科书，这是就意见提问。被调查者的记忆往往不清楚，或是故意给人一种特殊的印象，就事实提问不一定能提取关于事实的答案；被调查者往往根据社会最能接受的规范来表示意见，就意见提问也不一定能够提取坦诚的意见。

4. 问题和陈述语。我们可以就某些题目来直接提问，也可以向被调查人提出一个陈述，然后请他表示是赞成还是反对（是真还是假）。这种陈述也可以作为问题而使用，同样可提取信息。在态度调查中，我们往往都是提供一些陈述语，例如：

您认为应该加强口语训练？　　是　　否

或

口语训练应该减少。　　赞成　　反对

这两种格式在提取坦诚答案方面并没有什么不同，调查人选择什么格式决定于他希望什么回答方式。这在下面还要讨论。

5. 事先决定的问题还是按照主要问题的答案来回答的问题。有些问卷要求回答人完成每个项目；有些问卷则要求回答人按照其对一个主要问题的回答来决定是否要回答另一个问题，

例如主要问题是您是否念过大学，而另一个问题则是您对大学里哪一个科目最感兴趣。只有对第一个问题作出肯定的回答的人，才需要回答第二个问题。又如一个向学校校长提出的问题是您是否喜欢全国统一的教学大纲。如果校长回答喜欢，第二个问题是为什么喜欢。如果校长说不喜欢，第二个问题是为什么不喜欢。这两个问题是很相似的，但不完全相同。请注意，在问卷里的第二个问题不是简单地问“为什么?”，而是问“为什么喜欢（或不喜欢)?”，这是为了避免引起含混。按照一个问题的答案来回答的问题叫做取决于反应的问题（response-keyed questions)。

10.7.5.2. 应该怎样回答问题？

和提问方式一样，回答问题的方式也有多种。

1. 开放性反应。开放性反应也可称为非结构性反应，这是就反应而言的，不是指提问。这种反应允许被调查者随意用什么方式来回答。例如我们向学生提问，“您是否赞成举行四六级英语考试?”如果学生的回答是“赞成”，我们就可以继续提出一个开放性的问题，“为什么您赞成四六级英语考试?”当然我们也可以提出一个结构性的问题，例如举出5个理由，由学生选择一个答案。

下面是一些开放性反应的例子：

● 您为什么要进高等学校读书？

____

____

_______________________________。

● 您认为标准化考试的好处是：________________________

____

_______________________________________。

● 为什么您进了大学以后反而没有像在中学那样用功？

____

____

_______________________________。

由此可见，在开放性反应中，调查者除了把问题提出来，并提供一定回答的篇幅以外，并不控制被调查者的答案。这种方式的好处是被调查人可以自由回答问题，并不拘泥于赞成或反对。但是这种反应方式会给数据的量化带来很多问题。

2. 填充反应。填充反应是一种间乎结构性和非结构性反应之间的反应方式，它也要求被调查者产生一个反应，而不是选择一个反应。但是它把回答的范围限制在一两个单词和短语。例如，

● 您父亲的职业是什么？__________。
● 您在哪个学校念大学本科？__________。
● 您在中学念了几年英语？__________。

由此看来，填充反应和开放性反应的差别是程度上的不同。在填充反应方式里，我们要求被调查者报告事实性信息。所提的问题使答案限于几个词。

3. 表格式反应。这也是一种填充反应，但是结构性较强，因为被调查者必须把答案填到表格里。

表 10.15　用表格收集反应

最近一次工作的职称	所做的具体工作	顾主名称	年薪	日期	
				起	止

表格反应要求填进去的是数字、单词或短语（一般均为关于个人的事实），但是也可以用来反映一个量表上的程度（见下一节）。表格是一种组织复杂的反应的方式，有些反应不仅是单一的信息，而是包括了各种信息。但是它不算是一种独立的反应方式，因为表格所要求填进的反应也可以表示为填充反应或下面要介绍的量表反应。

4. 量表反应。最常见的结构性反应是量表反应，Tuckman 所提出的问卷（表 10.16）中的第一大项使用的就是量表反应。被调查人实际上必须在一个量表上作出选择，表示他对各种情况能否影响他工作提升的看法：

会使我停下来	可能会使我停下来	会使我认真考虑，但不会停下来	不会有任何影响

这是一个影响程度的 4 点量表，从有影响到完全没有影响。再看第三和第五大项，问的是达到目标的机会有多大。使用的是一个 5 点量表。也有一些量表是记录频率的，如：

很少发生	有时发生	经常发生	很经常发生

所有的量表反应都是测量频率或同意的程度，它们假定在量表上的一个反应是判断或感觉的量度。量表反应收集的都是可用的数据。在某些场合里，量表反应还可以看成是一个区间量表，例如在上述量表里，“很少发生”到“有时发生”的距离和“有时发生”到“经常发生”的距离是相同的。这些区间数据可以使用参数统计方法来分析。

5. 排列反应。如果我们给一系列陈述语，然后再要求回答者按照某一特定标准来把它们排列成次序，我们就能获得序数量度。例如：

按照它们对你学习怎样写行为目标的用途，把下列活动排列成次序。（从 1 到 5，1 表示最有用，5 表示最少用途。如果有那些活动完全没有用，就用 0 来表示。）

表 10.16 一份问卷的样本

Ⅰ. 假定你有一个机会，使工作和职务得到大大提升，请在下表中的每一个项目里对它会不会使你停下来，不考虑提升，作出选择，并划一个勾。

	会使我停下来	可能会使我停下来	会使我认真考虑，但不会停下来	不会有任何影响
影响你的健康				
使你有一段时间离开家庭				
经常在国内到处走动				
离开你的社区				
离开你的朋友				
放弃休息				
不能发表政见				
学习一种新的工作方式				
比你现在的工作更辛苦				
负有更重的责任				

Ⅱ. 从你现在的工作出发，你希望 5 年后会做什么工作？

____。

Ⅲ. 你达到上述目标的机会有多大？

____很大____大____一般____不很大____不大

Ⅳ. 你喜欢在 5 年后做什么工作？

____。

Ⅴ. 你达到上述目标的机会有多大？

____很大____大____一般____不很大____不大

________首先由顾问做介绍
________首先作小组活动
________每周开全体成员会议
________邮递与行为目标有关的指示和例子
________与顾问开个别会议

排列可以迫使回答人在不同备择中作出选择。如果要他们对上述活动评分，或表示同意或反对，他们有可能说成都重要。但要他们去排列，就要表示哪一个更重要。对排列数据的分析一般是把不同被试对同一项目的反应汇总起来，然后求出整个组的排列次序。我们可以使用非参数统计方法来比较一个组（如教师）的排列和另一个组（如教学管理人员）的排列。

6. 清单反应。在一份清单里，回答人只需要对所提供的选项作出选择。这种反应方式并非使用连续性的量表，而只是名义范畴，例如：

我最喜欢做的工作是：

选择一个选项

__________ (1) 我总是认为有能力完成的工作。

__________ (2) 我能尽量施展我的能力的工作。

清单要求作出的这些名义判断比较容易制定，回答人作答也不需要很多时间；但是所提供的信息较少。对名义性数据通常使用卡方检验来分析。在后面将专门讨论。

7. 范畴性反应。范畴性反应和清单反应相同，但更简单，只提出两个可能的选择，一般是是/否。例如，

你是否是一个老手？是________ 否________

有指导性的咨询开始得太晚了。真________ 假________

这些真假的数据也可以通过统计真实的反应的数目来变成区间数据，例如把所有被试认为真的反应加在一起以表示真实性的程度。

10.7.5.3. 建立问卷或访问提纲

1. 规定要测量的变量。在问卷和访问里要问的问题反映了我们所要了解的东西，例如我们的假设或研究课题。为了找出我们所要测量的东西，我们必须把研究中所有变量的名称写下来。例如我们要研究的是不同地区的学生学习外语的差异，我们就必须测量他们外语差异的情况以及他们来自什么地区。如果我们想了解本科生和大专生对就业的看法，我们就必须先让他们回答目前在念哪个年级，然后向他们提出一系列关于就业的问题。所以作为设计的第一步是把我们所要了解的变量规定得清清楚楚。

2. 选择提问的方式。是使用问卷还是使用直接访问？一般来说，问卷的方式比较经济和方便，但是在提问方面却有些限制，难以取得一些个人较敏感的、有启示性的答案，一些间接的、需要提示的问题也难以取得答案。而且问卷需要事先把问题都设计好，不能改变。下表比较在两种办法的差异：

表 10.17　　访问与问卷的比较

考　虑	访　问	问　卷
(1) 收集数据的人力	需要访问者	需要秘书
(2) 主要费用	雇用访问者	邮资和印刷费
(3) 根据反应再进一步提问题的机会 (个人化)	广泛	有局限
(4) 提问的机会	广泛	有局限
(5) 提示的机会	有可能	较困难
(6) 归纳数据的相对广度	大 (因为可进行编码)	限于花名册
(7) 所能找到的回答人数目	有局限	广泛
(8) 回报率	高	低
(9) 误差来源	访问者、工具、编码、样本	限于工具和样本
(10) 总体信度	有相当限制	较好
(11) 写作技能的要求	很少	很高

3. 选择反应的方式。选择哪一种反应方式并没有固定的规则。在某种情况下，你所要求的信息决定了哪一种反应方式较为适合；但有时你必须在两种方式中选择一个。例如你可以划一条线，让回答人填上自己的年龄；但你也可以提供一系列的年龄组，让回答人从中选择一个答案。选择哪一种反应方式决定于你是怎样处理所取得的数据；但不幸的是，我们有时在收集数据前并未决定应怎样处理数据。最好是把数据分析的决定和选择反应方式联系起来，使研究者能够 (1) 让收集到的数据服务于自己的目的；(2) 建立数据花名册，准备分析数据。例如关于年龄的数据最后将用于建立名义量表以做卡方检验，那么最好在问卷上提供一系列年龄组，让回答人作出选择。不同的反应方式会提供不同类型的数据，我们可看下表：

表 10.18　　不同反应方式的优缺点比较

反应方式	数据类型	优点	缺点
填充	名义性	较少偏颇，较大灵活性	较难评分
量表	区间性	容易打分	较花时间；可能有偏
排列	顺序性	容易打分，迫使作出区别	难于完成
清单或范畴	名义性 (汇总后可变为区间性)	容易打分；容易回答	提供的数据较少，选项不多

在选择反应方式时往往还有一些实际的考虑，如量表反应比真假选择题需要更多的时间来完成，假定你所设计的问卷已经太长了，那就不如用一些正误选择题。填充题的优点是不易偏颇，但却难以打分。取决于反应的问题有较大的灵活性，但却不易评分。

所以选择什么样的反应方式应考虑几个方面：

(1) 期望取得什么样的数据。如果期望的是区间数据，那就需要使用量表或清单反应。排列提供顺序数据，而有些清单只给名义性数据。

（2）反应的灵活性。填充题的灵活性最大，而是非题或正误题的灵活性最小。

（3）完成的时间。排列需要完成时间最长，量表反应也同样使人厌烦。

（4）潜在的反应偏颇。量表反应和清单反应都容易产生偏颇，除了回答人本身的一些社会价值观会导致偏颇外，还会有另外一些因素，如过多地使用“正确”和“是”，或在量表上固定一点，或是不愿选择两头。排列和填充则不会产生这种现象。排列还会迫使回答人作出区别。

（5）评分的难易。填充的答案还必须再编码，评分较困难。其他的反应都较易评分。

4. 准备访问项目。首先当然是规定要测量的变量，然后围绕这些变量设计问题。例如变量是学校气氛的开放性，最简单的问题是问教师。“您觉得这里的气氛有多开放?”而较为间接的问题是：您会不会随便把问题提到校长那里去？您会不会自由决定采取新的方法和教科书？下面是Tuckman设计的一个电话访问的提纲的一部分，目的是测量公众对公立学校的态度（如费用、质量、大纲、标准，等等）。为了取得最大限度的信息，这个提纲是设计得很严密的。

表10.19a　　一个电话访问提纲（部分）

现在我就New Jersey的公立学校提出几个问题：

21. 我们通常把学生评为A，B，C，D和不及格，以表示其学习质量。如果要您使用同一种评分办法来评估您的社区的公立学校，您评为A，B，C，D或是不及格？

135—1. A
2. B
3. C
4. D
5. 不及格
9. 不知道

22. 你觉得在您所在的社区里，花在公立学校的钱够不够多？

136—1. 是————→跳到第六页的第24题
2. 是
9. 不知道————→跳到第六页的第24题

如果对22题的回答为“不够多”，就问下列问题：

23. 为了要对公立学校花更多的钱，您是否愿意增加地方税收？
137—1. 是
2. 非
3. 不知道

表 10.19b　　一个电话提纲（部分）

24. 就目前社区花费在公立学校的情况而看，您觉得您社区的公民的钱是否值得这样花？
    138—1. 是
        2. 非
        3. 不知道
25. 您认为学生从高中毕业是否需要通过国家考试？
    139—1. 是
        2. 非
        3. 不知道
26. 您觉得学校应该在下列三个方面中强调哪一个方面为最重要的？[在指定地方开始，读下列选项]
    140—1. 数学和阅读教学
        2. 帮助学生升到大学或找职业
    或 3. 价值观和道德行为的教育
27. 哪一方面为次重要的？
    141—1. 数学和阅读教学
        2. 帮助学生升到大学或找职业
    或 3. 价值观和道德行为的教育

5. 准备问卷项目。准备问卷项目和准备访问项目是平行进行的。项目和变量的关系也是至关重要的，我们必须经常问自己，这是否我要测量的项目？这些问卷项目的形式在上面已介绍，这里就不重复。

6. 问卷的试点评估。在正式使用一份问卷之前，我们必须对它做试点评估。参加试点评估的人必须和将来使用问卷的人属于同一总体的，可以反映他们的意见的。如果我们围绕一个变量而设计了一组问题，这些问题应该都是测量同一变量的，因此我们可以求每个人对每个项目的评分和他对所有项目的评分的相关系数。相关大，就意味着这个项目所测量的正是所有项目所要测量的东西。例如我们发现多数的相关系数为 0.60 到 0.90 之间，而有几个项目在 0.10 到 0.30 之间，这几个项目就必须剔除。试点评估也可以帮助我们发现提问是否准确、清楚，不会引起误解。如果试点评估中所有的人对一个项目的回答都是一样的，这就说明这个项目没有什么区别性。如果有些回答人拒绝回答一些问题，这就说明这些问题十分敏感，我们必须改变一些字眼。

11. 统计方法

统计学(Statistics)来自 state+istics，指的是收集对国家极端重要的人口和经济信息。但是统计学已经发展成为在自然科学、人文科学和社会科学广泛应用的一种科学分析方法。在语言学研究中，描写语言学、应用语言学、实验心理语言学、语言习得、语言测试、社会语言学、计算语言学、语篇语言学等等，无不应用到统计方法。为语言学学生而编写的教科书也越来越多，如 Butler (1965)，Hatch & Farhady (1982)，Woods et al (1986)，Brown (1988) 等等，如果不掌握统计方法，语言学的研究者不但无法开展科学研究，甚至连别人的论文和报告也看不懂。

那么使用统计方法来做研究有些什么特征？Tuckman，Brown 等归结了几个方面：

- 系统性。统计方法具有严密的结构，必须遵循特定的程序性规则。怎样设计一项研究，怎样控制可能影响研究的各种因素，怎样选择和应用合适的统计方法，都有严格的规程。这些规程使我们的研究具有系统性，帮助我们了解、解释、批评各种统计研究报告。这些规程也成为统计方法研究的逻辑的基础。这不是说别的研究方法就没有系统性，而只是强调统计方法的系统性。
- 逻辑性。这些规程构成了一种直截了当的逻辑的型式——每一个构件都按部就班地推移到另一个构件；从逻辑的发展来看，每一个构件都不可缺少。如果规程受到破坏，一些构件就会丢失，逻辑就要中断。从课题的选择、抽样、提出假设、组织试验以验证假设、一直到解释数据和取得结论，都必须按规程严格执行。
- 可触摸性。统计方法从现实世界中收集数据，这些数据表现为分数、次序排列、频数，都是可以看得见、摸得着的。数据的类型虽然很多，但是它们都是量化的。这些数据的处理使我们的研究和客观世界挂起钩来。当然这并非说别的研究就不可触摸，而只是强调统计方法的这种可触摸的特性。
- 可重复性。用统计方法进行的研究都是可以重复的：研究者提出问题的方法和逻辑，所设计的实验方式，所收集数据的手段，别人都可以重复和验证（即在相同的条件下再做一遍）。可重复性是一个衡量研究质量高低尺度。
- 可简约化性。统计方法可以使日常生活中的一些语言和语言教学的复杂的问题简约化，使我们了解因素之间的相互关系，发现其基本的模式。

当然统计方法也不是万能的，它也有本身的问题，例如现实生活很复杂，许多因素纠缠在一起，并非那么容易控制这些因素，而且统计方法的使用也有是否得当的问题。使用不当往往会导致错误的结论。但只要是和数字打交道，就不可能离开统计方法；而且不使用统计方法，也不见得就能解决使用统计方法所带来的问题（如控制因素）。

研究语言和语言教学的人往往对统计学产生一种恐惧和畏难心理，以为它要求一些高深的数学知识，其实只要具备普通的算术常识，也能掌握其基本的原理和算法。有鉴及此，不少的统计学书籍都故意避开高等数学。随着计算机的普及，已有很多统计软件包问世，而且越来越便于操作，所以作统计运算，根本就不成问题。主要的问题是掌握其基本的原理，知

道在什么时候使用什么统计程序，正确地使用统计结果来解释所研究的问题。在以下的章节里，我们将着重解释一些基本的统计学概念，讨论它们的用途，同时介绍一些统计软件包（如 SPSS6.0，STATISTICA5.0，EXCEL5.0）的使用方法。我们也介绍一些基本的公式，但不作太多的说明，更不去做数学的推导。

11.1. 描写统计方法

统计方法有两大类：一是描写统计方法（descriptive statistics），又叫归纳统计方法（summary statistics），其目的是通过有关的量度来描写和归纳数据。二是推断统计方法（inferential statistics），其目的是根据对一小部分数据的观察来概括它所代表的总体的特征。科学考察都是为了概括更多的事实，故需要用到推断统计方法。例如我们对一个小班的学生教英语，在期末举行了一次考试，对考试结果作统计分析，它告诉我们，这个班的总成绩如何，每一个学生的成绩如何，它和别的班相比的成绩又如何等等。我们关心的是这个班的学习成绩，采用了描写统计方法（如平均分、标准差、分布情况等等）。又假定我们想试验一种新的教学方法，选了一个小班来作实验，经过一个学期的教学以后，我们想比较这个班的成绩和没有用新教学方法的另一个班的成绩，看有没有明显的差异。对这两个班的成绩，我们首先都要进行描写，这是比较的基础。比较的结果是用新教学方法的那个班明显地比没有用新教学方法的那个班好，但是我们感兴趣的是通过它们成绩的比较，我们能否推断（1）这个班的成绩比另一个班的成绩好是因为使用了新教学方法？（2）如果以后还要在另一个班里使用这个方法，我们有多大的把握使其教学效果高于不使用这个方法的班？这就需要到推断统计方法来概括新方法是否比旧方法好。

由此可见，使用描写统计方法的不一定要使用推断统计方法，但使用推断统计方法的要首先使用描写统计方法。

11.1.1. 数据的归纳

在上一章，我们已经介绍过范畴性（名义性）数据和区间数据的区别，现在进一步来看这两种数据应怎样进行归纳。

11.1.1.1. 列表

对范畴性数据的归纳一般采用交叉列表（cross-tabulation）的方法。这种方法很简单，先把每个范畴的频数列出，然后计算其百分比。例如有人在美国调查了 560 个有语言缺陷的人，其中 364 人是男性，196 人是女性。这些语言缺陷分为四类：口吃、语音缺陷、特殊语言紊乱、听觉损伤。就可以按 2×4 个范畴来交叉列表：

表 11.1　　560 人四种语言缺陷按性别列表（按总体的相对频率计算百分比）

	口吃	语音缺陷	特殊语言紊乱	听觉损伤	总计
男性	57 (10)	209 (37)	47 (8)	51 (9)	364 (65)
女性	27 (5)	118 (21)	31 (6)	20 (4)	196 (35)
总计	84 (15)	327 (58)	78 (14)	71 (13)	560 (100)

在括号里的是总体的百分比，亦称相对频率（relative frequencies）。我们也可以按照需要，以行（性别）或列（语言障碍）为单位，计算百分比，如：

按性别（行）总数的相对频率来计算：

表 11.2　　560 人四种语言缺陷按性别列表（按行总数的相对频率计算百分比）

	口吃	语音缺陷	特殊语言紊乱	听觉损伤	总计
男性	57 (16)	209 (57)	47 (13)	51 (14)	364 (100)
女性	27 (14)	118 (60)	31 (16)	20 (10)	196 (100)
总计	84 (15)	327 (58)	78 (14)	71 (13)	560 (100)

按语言障碍（列）总数的相对频率计算：

表 11.3　　560 人四种语言缺陷按性别列表（按列总数的相对频率计算百分比）

	口吃	语音缺陷	特殊语言紊乱	听觉损伤	总计
男性	57 (68)	209 (64)	47 (60)	51 (72)	364 (65)
女性	27 (32)	118 (36)	31 (40)	20 (38)	196 (35)
总计	84 (100)	327 (100)	78 (100)	71 (100)	560 (100)

这些表格可用图来表示，图 11.1 是一个条形图。

交叉列表主要是用来表示一些语言调查和问卷调查或访问所收集的数据，这是一种基本的描写数据的方法。通过表格和图表，往往也能揭示一些事物的型式，例如在语言障碍调查方面，我们可以明显地看到，语音缺陷是主要的语言障碍型式，不管男性还是女性都是如此。

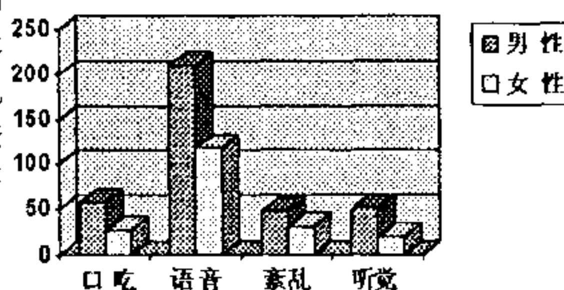


图 11.1　语言缺陷条形图

交叉列表的制作很简单，可以人工制表，在统计软件包（如 SPSS，STATISTICA）中也很易操作。在这些软件包里有一个表格（spreadsheet），可把数据打进去，但要注意格式。交叉列表的格式如下：

Var 1	Var 2	Var 3
1	1	57
1	2	209
1	3	47
1	4	51
2	1	27
2	2	118
2	3	31
2	4	20

第一行的 Var 是表格提供的，可以重新定义为字符串。在其他的一、二两列里，我们用数字来代表字符串（第一列中 1＝男性，2＝女性；第二列中 1＝口吃，2＝语音缺陷，3＝特殊语言紊乱，4＝听力损伤），也可以直接输入字符串。如果输入的是汉字则要求有一个有汉字的工作平台，如 Windows 3.2 或 Windows 95 的大陆版。要注意的是第一和第二列是交叉列表的符号，而第三列才是实际的频数。在 SPSS 里，我们必须在 Data（数据）的菜单里，选择 Weight Cases（对案例加权），然后把第三列定义为 frequency（频数）。在 STATISTICA，也需要在主菜单的 Weight（加权）定义频数。然后再在 SPSS 的主菜单里去 statistics/summarizes/crosstabs[①]，或在 STATISTICA 的主菜单里去 Basic Statistics/Tables and Banners/crosstabulation。

11.1.1.2. 频数表

如果我们所收集的是区间数据，即连续性数据，例如一个考试的分数，一个实验的反应时等等，那就要用另一种归纳方法——频数表（frequency table）。假定 100 人参加一个考试后的成绩为：

97	83	72	67	63	56	42	95	78	72	66	62	55	41	57	68	65
94	78	71	66	62	55	40	94	78	71	66	60	54	40	42	58	45
93	77	71	66	60	50	34	92	77	70	66	60	50	34	64	85	74
91	77	70	65	60	49	33	91	77	69	65	60	49	30	67	67	64
90	77	68	65	60	48	30	90	76	68	65	59	47	30	73	57	
90	76	68	65	59	46	88	75	68	65	59	46	87	75	84	44	

对这些分数的整理可以采取几个步骤：

(1) 决定区间，如以 10 分作为一个区间。区间的大小是任意决定的，但一般不要小于 5 个，大于 15 个。

(2) 把分数排序。

(3) 按照区间算出分数的频数，如 20 到 30 这个区间，共有三个 30，30，30，其频数就是 3。

(4) 计算其累计频数。

(5) 计算相对频数，将每一区间的频数除以总数（这组数据刚好是 100），相对频数也就是比例。也可以表示为百分比，即用 100 乘比例。例如 0.05 为相对频数（比例），而百分比则为 5%。

---

① 我们用符号/来表示路径，即从主菜单去 statistics 菜单，再去 summarizes 菜单，最后选 crosstabs。到 crosstabs 后，再用鼠标双击框图里的图标进行选择。下同。

（6）计算累计相对频数。

表 11.4　　100 人的分数频数表

区间	组中值	频数	累计频数	相对频数	累计相对频数
21～30	25.5	3	3	0.03	0.03
31～40	35.5	5	8	0.05	0.08
41～50	45.5	13	21	0.13	0.21
51～60	55.5	16	37	0.16	0.37
61～70	65.5	28	65	0.19	0.65
71～80	75.5	19	84	0.19	0.84
81～90	85.5	8	92	0.08	0.92
91～100	95.5	8	100	0.08	1.00

累计相对频数表示的是百分位，如 80～90 那个区间的累计相对频数是 0.92，这意味着这个分数档在 100 考生中高于 92%的考生，低于的 8%的考生。频数表可以表示为直方图，图中的曲线为期望的分布曲线，可以帮助我们了解这组数据的分布是否正态。

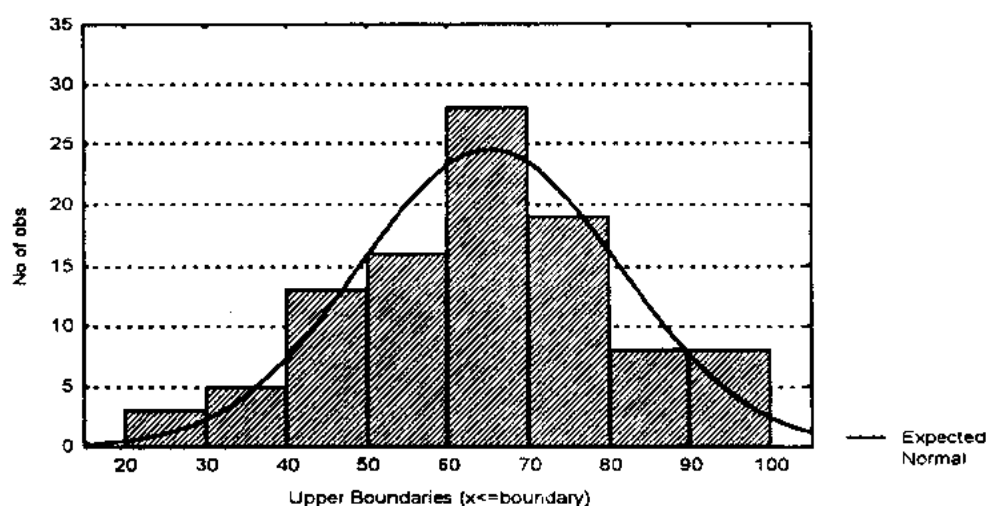


图 11.2　频数排列直方图

在 SPSS 和 STATISTICA 里只需把数据排成一列输入（程序会自动排序），然后在 SPSS 主菜单去 statistics/summarize/frequencies，在 STATISTICA 可从主菜单直接选择 frequency tables，即可按照菜单指令计算结果和作图。

11.1.2. 集中量

集中量（Measures of Central Tendency）是描写统计方法最常用的一种度量。有时我们面临一大堆数据，而逐个罗列这些数据又费时失事，往往需要用一个数字来描写数据，集中量是最合适的表示方法，因为它表示的是数据的趋中位置。集中量又好几种，最常用的有平

均数（mean），中位数（median）和众数（mode）。

11.1.2.1. 平均数

平均数是算术平均数（the arithmetic mean）的简称，可定义为所有观察值之和除以所有观察数（N）：

$$\bar{X}=\frac{\sum_{i=1}^{N}X_i}{N} \quad (11.1)$$

但是一般的频数表的数据都比较多，逐个累加也比较麻烦，所以可利用频数表的区间算出组中值，以表 11.4 为例，每个区间为 10 分，因此组中值应为第 5 个值加第 6 个值除以 2，故 21～30 那个区间的组中值为（25+26）/2 = 25.5。用频数表来计算平均数的公式是：

$$\bar{x}=\frac{\sum fx}{N} \quad (11.2)$$

f 为频数，x 为每个区间的组中值，N 是总数。

以表 11.4 为例，平均数应为（25.5×3 +35.5×5+45.5×13……+95×8）/100=64.5。用频数表算出的平均值，有时和逐个计算会有些细微差别，这是因为它使用了区间，精确性受到影响。例如表 11.4 的真正的算术平均值为 65.18。由于统计软件包的普及，不需要用手算，故也无须考虑使用频数表或逐个计算。在 SPSS 里，我们可以在计算出频数表后，从主菜单去 statistics/summarize/descriptives ，便可获得包括集中量在内的各种描写统计量。在 STATISTICA，则去 basic statistics/descriptive statistics。

11.1.2.2. 中位数

中位数也是一种集中量，它的计算方法是按顺序把数据排列，取其中央位置的值，例如 2，4，5，5，7 的中位数为 5。假定数组不是奇数，而是偶数，那就取其中间的两个数据的算术平均数，如 2，3，5，6，7，7 的中位数为（5+6）/2=5.5。所以实际上，中位数就是刚好处于 50%的百分位的那个值。中位数值和算术平均数可以刚好是一致的，但在大多数的情况下，略有一点差异。以表 11.4 那组数据为例，算术平均数为 65.18，而中位数则为 66。

11.1.2.3. 众数

众数是一组数组中频数最多的那个数值。众数可用简单的目击的方法求出。在 2，4，5，5，7 的数组里，众数是 5，中位数是 5，而其平均数为 4.6，三者比较接近。但在 2，3，5，6，7，7 的数组里，众数是 7，而中位数是 5.5，平均数为 5，中位数和平均数较接近，但众数和两者相差较大。由此看来，众数比较粗略，而且和数组的分布有关。有时有些数组有两个甚至更多的数值的频数刚好是一样，众数也就求不出来。

11.1.2.4. 几个集中量的比较

1. 算术平均数、中位数和众数都是集中量，按理应该统一起来，实际上，如果频数的分布是正态的，这三者是统一的。

这一个数组的算术平均数、中位数和众数均为30，这是因为数组的分布是正态的，如图11.3。

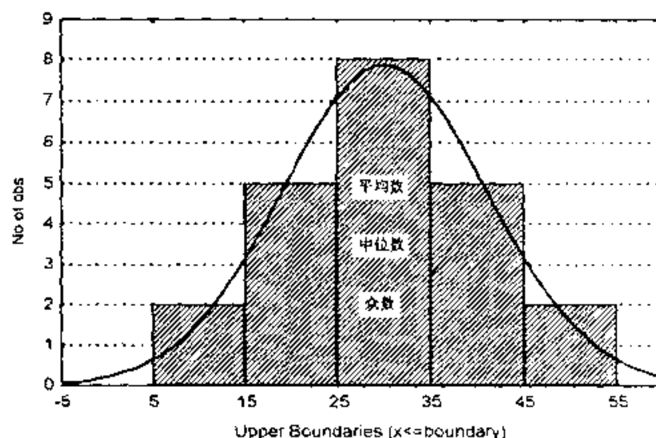


图11.3　平均数、中位数和众数一致：正态分布

表11.5

分数	频数
10	2
20	5
30	8
40	5
50	2

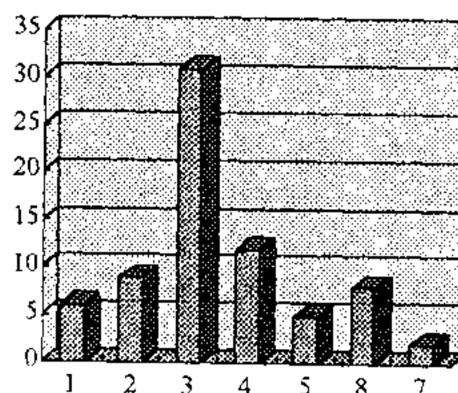


图11.4　众数比平均数更有代表性

2. 如果出现偏态时，这三者就不会一致。在正偏态时，平均数＞中位数；在负偏态时，平均数＜中位数。图11.2略呈负偏态，而其平均数位65.18，中位数为66。

3. 算术平均数考虑到每一个数据，感应最灵敏，不大容易受到抽样变动的影响，最能代表数组。而且它的计算方法简单，容易操作，故一般均使用它来表示数组的趋中倾向。但是，正因为它考虑了每一个数据，所以常常会受到一些极端的数据的影响，反而不如中位数那样稳健robust)。例如2，3，5，6，7，20这一个数组，平均数为7.2，可是中位数却只有5.5。这个平均数显然是受到一个极端值20所影响，如果我们把它去掉，平均数只有4.6。在这种情况下，平均数反而不如中位数稳健。有些时候，平均数又不如众数。在图11.4里，众数是3，显然最有代表性，倒是平均数被大于3的几个值拉高了，变成3.42。

11.1.2.5. 修剪平均数

上面谈过，有时平均数受极端的分数影响，反而不如中位数，这种情况在心理语言学实验中对反应时的统计常有出现：有的被试精神不集中，反应时往往比别人长很多；有的被试敷衍了事，反应时又会特别快。这时中位数反而更真实，但是用中位数来计算t－检验，却有些麻烦。于是就提出修剪平均数（the trimmed mean）的问题。在一些统计软件包里，我们可以

自己定义修剪两头的极端分数，如 5%、10%等等。这往往又有点任意性。

我们不妨先看一个包括 12 个值的数组：

81，94，95，95，97，102，106，110，113，114，135，160

我们可作不同程度的修剪：

a. 0%

b. 8%

c. 25%

d. 42%

a. 零修剪就是原来的平均数，(81+94+…+135+160) /12=108.5

b. 修剪 12 个值的 8%相当于从每一端减去 1 个值，(94+…+135) /10=106.1

c. 修剪 12 个值的 25%相当于每端减去 3 个值，(95+97+102+106+110+113) /6=103.6

d. 修剪 12 个值的 42%相当于每端减去 5 个值，这就是中位数 (102+106) /2=104

这个数组的中央 (m) 是 100，160 明显地影响了平均数的估算，修剪有所改善。但如果一直修剪到 42%，却反而不那么准确了。如果我们仔细看 25%的修剪，它最接近 100，但共减去了 6 个值，即一半，又不免觉得去掉的信息太多。如果数组为 10 000，有 5 000 个数值被删去。于是有些统计学家 (Mosteller & Tukey，1977，见 Wonnacott & Wonnacott) 觉得不一定都删去，而是应该逐步减少权重，越是离异数据 (outliers)，权重就越小。对一个观察值 X 的权重的公式为：

$$w(X) = (1-Z^2)^2 \quad If|Z|\leqslant 1$$
$$=0 \quad If|Z|\geqslant 1 \tag{11.3}$$

Z 是按照中位数和四分位区间距 (interquartile range，简称 IQR) 来对 X 值作标准化处理。其公式为：

$$Z=\frac{X-X}{3\ (IQR)}, \tag{11.4}$$

X 为中位数。最后再求出双权重平均数 (biweighted mean)，其公式是：

$$Biweghtedmean \quad \overline{X}_b=\frac{\sum Xw\ (X)}{\sum w\ (X)} \tag{11.5}$$

下面我们以这个数组为例来说明，这个数组的第一个四分位 (即 25%的位置) 为 95，第三个四分位 (即 75%的位置) 为 113，IRQ 为 113−95=18。

表 11.6　**双权重平均数的计算方法**

分数 (*X*)	*Z*	*w* (*X*)	*Xw* (*X*)	
81	−0.42593	0.670085	54.27687	
94	−0.18519	0.932589	87.66336	
95	−0.16667	0.945216	89.79552	
95	−0.16667	0.945216	89.79552	
97	−0.12963	0.966675	93.76744	
102	−0.03704	0.997258	101.7204	中位数＝104
106	0.037037	0.997258	105.7094	第一个四分位＝95
110	0.111111	0.975461	107.3007	第三个四分位＝113
113	0.166667	0.945216	106.8094	四分位区间距＝18
114	0.185185	0.932589	106.3151	
135	0.574074	0.449488	60.68093	
160	1.037037	0	0	
总数		9.757052	1003.835	
双权重平均数＝		102.883		

从表中看，只有 160 这个值是大于 1 的，这是一个离异数，应该剔除。用（11.5）计算，双权重平均数为 102.883，比直接修剪 25%的数据又有所改善。我们也可以用一种迭代法来改善双权重平均数，其结果为 102.627。具体的运算这里就不介绍了。

11.1.3. 离散量

离散量（Measures of Dispersion）描写数据的差异程度和离散程度。光使用集中量来描写一个数组，往往是不够的，例如下列两个数组：

表 11.6　**两个平均数相同，但离散程度不同的数组**

	1	2	3	4	5	6	7	8	9	平均数	全距
A 组	50	59	60	61	62	63	63	68	72	62	22
B 组	28	35	55	60	65	70	75	80	90	62	62

这两组学生成绩的平均数都是 62 分，但是他们的离散程度很不一样，A 组学生显然不如 B 组学生的分布广。如果我们不对平均数辅以分布情况的说明，就简单认为这两组学生的水平是一样的，就会引起一些误解。

11.1.3.1. 全距

对数据的离散程度的最简单的说明是全距（Range），即最高分到最低分的距离。A 组的全距为 72－50＝22，而 B 组的全距为 90－28＝62。全距的计算办法简单，而且很容易理解，但是它只用到两端的数值，没有考虑中间数值的差异，所以感应不灵敏。另一种表示距离的方法就是上面谈到的四分位区间距（IQR），即用第三个四分位的值减去第一个四分位的值，在

A 组为 63－60＝3，在 B 组为 75－55＝20。四分位区间距虽然比全距好些，但它删去 50％的数值，最好的办法是把全部数据都考虑在内来计算分布程度。

11.1.3.2. 方差与标准差

方差（variance）与标准差（standard deviation）把每一个数值都考虑在内来表示数据的离散程度。假定我们有一个数组，X_1，X_2，…，X_n，已知其平均数，我们就可以计算每一个数值和平均数之差。那么这些差的平均数不就可以反映这些数值和平均数之间的距离了吗?但是实际执行的结果是这些差有正值，也有负值，而且它们加起来刚好等于 0（见表 11.7 中的 2）。所以我们必须把这些差乘方后再累加起来，然后再除以数值的个数 n－1，这就是方差（s^2）。方差的开方就是标准差，或称均方差（s）。

$$s^2=\frac{\sum (X-\overline{X})}{n-1}, \qquad (11.6)$$

$$s=\sqrt{s^2} \qquad (11.7)$$

如果从频数表计算，则标准差的公式为：

$$s=\sqrt{\frac{\sum (X-\overline{X})^2 f}{n-1}} \qquad (11.8)$$

具体的运算请看表 11.7：

表 11.7　　标准差计算方法

	1：X	2：$X-\overline{X}_2$	3：di
1	28	－34	1 156
2	35	－27	729
3	55	－7	49
4	60	－2	4
5	65	3	9
6	70	8	64
7	75	13	169
8	80	18	324
9	90	28	784
总数		0	3 288
平均数（$\overline{X}$）	62		
方差（s2）	＝3288/8	＝411	
标准差（s）	20.27313		

但是这种计算方法需要先求出平均数，再求离差，多了一个循环。若在计算机编程，则不如下一个公式直截了当：

$$s=\sqrt{\frac{\sum x_i^2-(\sum x_i)^2/n}{n-1}} \qquad (11.9)$$

$\sum x_i^2$ 称为平方和，而$(\sum x_i)^2$ 称为和的平方，这两个概念容易混淆，但在以后的方差分析运算中常常碰到，必须弄清楚其差异：一个是先把每个数值平方，然后再累加；另一个是先把每个数值累加，然后再平方。

11.1.3.3. 偏态值与峰值

偏态值（skewness）与峰值（kurtosis）都和正态分布有关，正态分布是两头小，中间大，平均数的两边是对称的，平均数、中位数和众数都是一致的，见图 11.3。但是有时两头是不对称的，就出现偏态，有正偏态和负偏态两种。

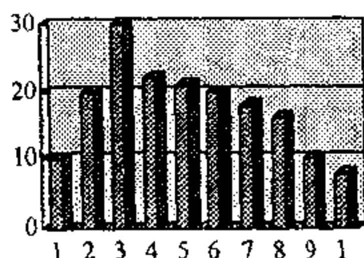


图 11.5

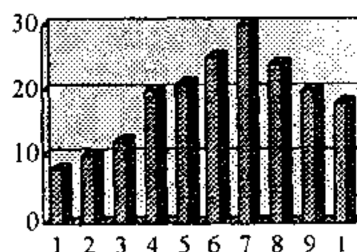


图 11.6

正偏态向左偏移，说明结果偏难；副偏态向右偏移，说明结果偏易。除了用目击法来判断偏态外，一般还可以使用公式，Pearson 的偏态指数公式为：

$$SK=\frac{3\ (\bar{x}-\dot{x})}{s} \tag{11.10}$$

$\bar{x}$ 为平均值，而 $\dot{x}$ 为中位数，以表 11.4 的数组为例，平均数为 65.18，中位数为 66，标准差为 16.23，故 SK=3（65.18−66）/16.23=−0.1515。另一个常用的公式是：

$$g_1=\frac{(\sqrt{n})\ \sum\ (X-\overline{X})^2}{[\sum\ (x-\bar{x})^2]^{2/3}}, \tag{11.11}$$

这个公式较为麻烦一点，但较精确，一般统计软件包采用它，结果为 −0.1671，和 Pearson 的偏态指数相差不大。如果分布是正态的、对称的，偏态指数是 0，负值表示负偏态，正值表示正偏态。

另一个和正态分布有关的概念是峰值，峰值表示的是分布曲线的最高点和正态的峰相差的程度，是过尖（称为 leptokurtic）？还是过平（称为 platykurtic）？计算峰值（g_2）的公式是：

$$g_2=\frac{n\sum\ (x-\bar{x})^4}{(\sum\ (x-\bar{x})^2)^2}-3 \tag{11.12}$$

如果所得值为负值，表示曲线分布过平，正值则表示曲线分布过尖。0 值或接近 0 值则表示正态或接近正态。以这个数组为例，峰值为−0.33，说明峰值比较平坦，各个分数段的频数差不多。

11.1.4. 描写统计方法的计算机运算

上面所谈的频数统计、集中量、离散量、偏态值与峰值在一些统计软件包里都是最基本的方法，可一次完成。例如在 SPSS 里，如果要排列频数表，可去 statistics/summarize / frequencies；如果要其他统计量，可以去 statistics/ summarize / descriptives 进入菜单，再作选择。图 11.7 是 SPSS 的结果报告：

STATISTICA 的操作办法大同小异，选择统计量的方法相同，但结果是分别显示，不是列在一个表上。EXCEL 则可以去 fx（函数），再选择“统计”下面的有关指令，如 sum（总计），average（平均值），median（中位数），mode（众数），stdev（标准差），var（方差），max

(最高值), min (最低值), skew (偏态值), kurt (峰值), standardize (标准分), 等等。SPSS, EXCEL 都可以作图, 但作图功能最强的是 STATISTICA, STATGRAPHICS。

15 Jun 96 SPSS for MS WINDOWS Release 6.0 Page 1

Number of valid observations (listwise) = 100.00

Variable VAR00001

Mean	65.180	Std Dev	16.235
Variance	263.563	Kurtosis	−0.330
S.E. Kurt	0.478	Skewness	−0.167
S.E. Skew	0.241	Range	67.000
Minimum	30.00	Maximum	97.00

Valid observations − 100 Missing observations − 0

图 11.7

11.2. 概率分布

上面谈到的是数据的描写,但是我们的研究兴趣往往不仅是了解一个数组的统计量本身,我们还想产生这个数组的样本在多大程度上能够代表总体(或称母体),即我们有多大把握可以把结果推论到总体。假定我们严格按照实验方法(包括随机抽样和使用控制组),在 30 个大学一年级的学生身上试验使用一种新英语教材,经过一年的试验后取得较好的成绩。我们并不满足于描写这 30 个人的成绩,我们更想知道,别的大学一年级学生使用新教材是否也会取得同样好的成绩。当然两次实验的结果不会是完全一样的,但是我们能否根据样本来作出一些推断:两次的样本是否属于同一个总体的?两次试验结果是否都能说明新教材是有效的等等。

要解决这个问题就需要使用到推断统计方法。这些统计方法以概率原理为基础。概率是一个难以定义的概念。但在现实生活里,我们经常都用它来谈论事物,例如小李的英语学得不错,让他参加 TOEFL 考试,他有 70%的机会可以拿 600 分;广州属亚热带气候,不大可能会下雪。实际上我们都是在根据已有的信息来作预测,我们掌握的信息越充分,预测的把握就越大。在推断统计方法里,我们想知道的是我们作出正确推断的可能性有多大。我们知道,如果我们不断地重复测量一些人类行为,其结果就会接近正态的分布。所以如果我们在一年级学生身上不断地重复上述的教材实验,我们所得的结果也会接近正态分布。

概率可以表示为期望的结果数除以可能的结果数,例如一个钱币有两面(即两种可能的结果),不是正面就是反面,所以期望的正面的概率就是 1/2= 0.50。这就是说,得到正面的机会是 50%。如果我们只掷几次,正面和反面的次数不一定刚好就是各占 50%,但要是我们不断地掷下去,结果就会越来越接近这个百分比。

正面的次数	4	53	139	237	480
掷的次数	10	100	300	500	1000
概率	0.40	0.53	0.46	0.47	0.48

这里需要区分表示概率的两种方式，一种是比例（proportion），即上面谈到的相对频数，表示为小数点后的数字，如0.40、0.53等；另一种是百分比（percentage），表示为数字后加%，如40%、53%等。用100来乘比例就可得到百分比。所以在一个试验里，概率就是我们不断重复该实验后所得到的相同结果的次数的比例。公式11.13的意思是，当n趋于无限大的极限值时，概率为期望结果数（m）除以可能的结果数（n）。

$$p = \lim_{n \to \infty} \frac{m}{n} \tag{11.13}$$

为什么我们要先介绍概率？原因有3个：

1. 我们必须认识某一事件的概率表示为期望的结果和各种可能的结果总数的比例。这个比例从0到1，表示为百分比就是从0%到100%。

2. 更重要的是，如果我们要作出决策，我们应该选择概率高的那一种可能性。

3. 我们需要应用概率的概念来讨论事件的型式和各种分布类型。

11.2.1. 离散性概率分布

表11.8　　样本中的男人的概率分布

随机变量的数值	概率
0	1/64
1	6/64
2	15/64
3	20/64
4	15/64
5	6/64
6以上	1/64
总计	1

11.2.1.1. 随机变量和离散性概率分布

和概率分布有关的另一个重要概念是随机变量（random variables）。随机变量是一个其数值取决于实验或观察的结果的变量，这也就是说，随机变量所具有的数值是和一个特定实验的样本空间的随机事件相联系的。例如我们从一个男、女性数目相同的总体中随机地随机抽6个人，6个都是男人或6个都是女人的概率都是比较低的，最大可能的是3个男人，或3个女人，从表11.8可见，概率为0.3125。从0到6是作为实验的结果而出现的，都是随机变量的数值。所以一个随机变量的分布就是这个变量的各个数值的概率的罗列。如果这个变量的不同的数值都能列出来，这个随机变量就是离散的随机变量。离散变量可以是数字（像这里所举的例子），也可以是范畴，如“性别”、颜色（如黑和白）。应该注意的是，一旦我们计算出概率后，总体的体积就不再是很重要的数字了。

为了进一步说明离散性的随机变量和概率分布的关系，我们可以从一个最简单的例子说

起。假定我们掷两次钱币，第一次不是正面就是反面，第二次也是一样。这样我们就有 4 种排列，这些排列的概率和它们的直方图可以表示如下：

表 11.9　　　　　　　　　　掷两次钱币的概率及其直方图

第一次	第二次		事件	频数	概率	直方图
	正	正正	2 正	1	1∶4 (0.25)	
正						
	反	正反				
			1 正	2	1∶2 (0.50)	×
	正	正反				
反						×　×　×
	反	反反	0 正	1	1∶4 (0.25)	1 正　2 正　0 正
				4		

这就是说，如果我们掷两次钱币，有 50%的机会是一正一反，而得到两正或两反的机会则相等，各为 25%。如果我们继续掷下去，三次、四次、五次……，这个变量的数值和概率都可以列出来的。例如掷 5 次的结果是：

事件	频数	概率
5 正	1	1∶32 (0.03125)
4 正	5	5∶32 (0.15625)
3 正	10	10∶32 (0.3125)
2 正	10	10∶32 (0.3125)
1 正	5	5∶32 (0.15625)
0 正	1	1∶32 (0.03125)
总计	32	(1.0000)

我们可以看到，要 5 个都是正面或反面的情况是非常小的，只有 0.03125，即不到 1/20。如果出现这样的情况，那就很可能这个钱币有偏。换句话说，要想了解一个钱币是否有偏，我们必须起码掷 5 次。实验中的显著性水平就是从此而来的，如果掷 5 次，5 次都是正面，我们就可以说这个钱币不同于一般的钱币，具有 0.03125 的显著意义（小于 0.05）。在实验中，我们往往要求达到这个水平，才能说结果在 5%水平上具有显著意义。

11.2.1.2. 二项分布

二项分布是一种离散型的概率分布。二项式定理根据组合的知识来研究二项展开式。组合的公式表示为：

$$C_n^m=\frac{n!}{m!\ (n-m)!},\qquad (11.14)$$

这个公式的意思是从 n 个元素中选出 m 个的组合，“!”表示阶乘。例如在上面掷 5 次钱币中，

要得到 3 正的频数为：

$$C_5^3=\frac{5\times4\times3\times2\times1}{(3\times2\times1)\ (2\times1)}=\frac{120}{12}=10$$

二项展开式可以概括成为一个通式：

$$P_{(x)}=C_n^x p^x q^{n-x}, \qquad (11.15)$$

表 11.10　p 为 0.25 和 0.5 的二项分布

	p=0.25	p=0.5
0	0.056314	0.000977
1	0.187712	0.009766
2	0.281568	0.043945
3	0.250282	0.117188
4	0.145998	0.205078
5	0.058399	0.246094
6	0.016222	0.205078
7	0.00309	0.117188
8	0.000386	0.043945
9	2.86E-05	0.009766
10	9.54E-07	0.000977

在上例，一个钱币只有两面，所以 p=0.50，而 q=1−p=0.50。所以掷 5 次钱币得到 3 次正面的概率为：

$$P_{(x)}=10\times0.503\times0.502=0.3125$$

二项分布的特点是，只要 p 不是 0 或 1，n 越大就越接近正态分布；如果 p=0.50，二项分布和正态分布是一样的。下面的表 11.11 比较了 10 个题目中答对题数的概率。一个的 p=0.25（如四选一的选择题），一个的 p=0.50（如正误题）。如果这 10 个题目是四选一的选择题，而学生又全然不懂，他猜对 5 题的概率是 0.058。但如果是正误题，猜对 5 题的概率是 0.246。一般的考试当然不会只有 10 道选择题，假定有 50 题，用二项分布的公式计算，靠猜测来答对一半题目的概率只有 0.0000845，接近于 0。

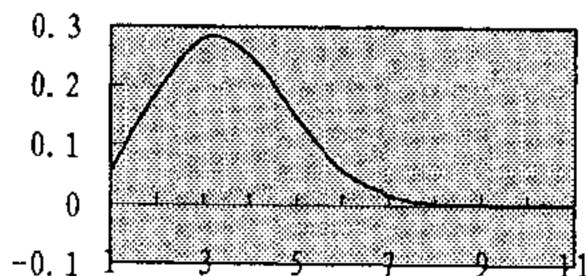


图 11.8　10 个项目（P=0.25）的二项分布

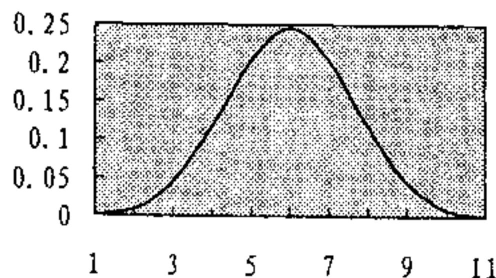


图 11.9　10 个项目（P=0.5）的二项分布

11.2.2. 连续性概率分布

11.2.2.1. 连续性随机变量

连续性概率分布在一个连续性样本空间的区间中取值，从理论上看，这些数值可以是无限的。假定我们要记录被试在一个实验中的反应时间，而我们的记录工具是很准确的，可以精确到百万分之一秒，我们的准确度是无限的。一个被试的反应时间可以是 1.643217 秒，但是我们还可以使之更精确，如 1.6432171 秒，甚至 1.64321705 秒，1.643217049 秒等等。具有这种特性的变量叫做连续性变量，而连续性概率就是反映这类变量的分布的。

如果一个变量的数据很多，而且组距又很密，我们可以根据数据的频数作一个直方图，并用一条曲线把直方图每个矩形上面的中点连接起来，所作出的就是一条平滑的曲线（见图 11.2）。这条曲线就是概率密度曲线（probability density curve），简称为密度或概率分布；它表示的是相对频数密度，或称概率密度函数。一个连续性随机变量 X 可以用 X 在 x 的概率密度来定义，写成 f（x）。如果 f（x）是连续性随机变量 X 的概率密度函数，那么 P（a<X<b）可用曲线 y = f（x）下的 a 到 b 的面积来表示。

一个随机变量的概率分布也可用积累概率 F（x）来表示，随机变量 X 取得小于或等于 a 值的概率可写成：

$$F(a) = P(X \leqslant a)$$

积累分布函数把 X 所有的数值和相应的积累概率联系起来。积累分布函数应该满足下列的数学特性：

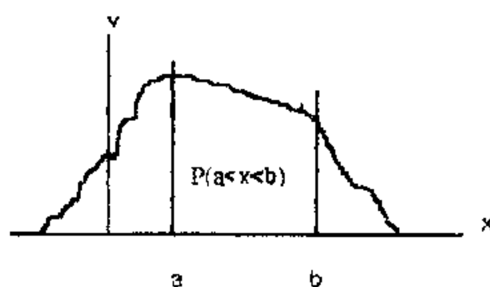


图 11.10　概率密度函数的图形

1. $0 \leqslant F(x) \leqslant 1$
2. 如果 $a < b$，那么 $F(a) \leqslant F(b)$
3. $F(\infty) = 1$ 而 $F(-\infty) = 0$

11.2.2.2. 正态分布

正态分布是连续性概率分布中使用得最广泛的一种分布。正态分布是对称的、钟形的（两头小、中间大）一种分布，它的数学模型最早由德国数学家 Gauss（1777—1855）提出，比利时数学家 Quetelet（1796—1874）对数据的收集有着深厚的兴趣，他进一步发现这个数学模型可用来解释很多社会学和人类学的现象。由于他在欧洲参加了 100 多个学会，使用正态分布的概念得以很快地在欧洲传播。例如 Quetelet 调查了 5738 名苏格兰士兵的胸围，发现它们和正态分布的频数十分一致①。胸围的平均数为 39.8 寸，标准差为 2.05 寸。

① Quetelet 原来公布的是观察到的频数和相对频数，然后按正态分布的公式推算的相对频数，我们根据这两个频数作成图，以提高直观性。

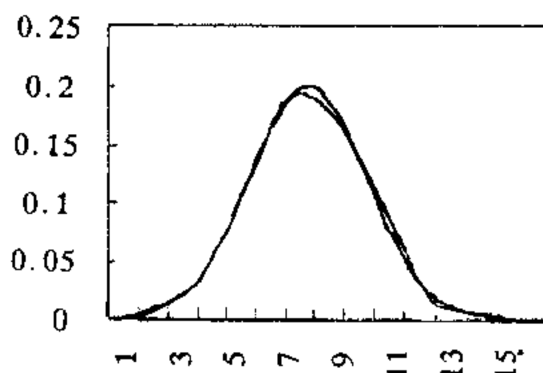


图 11.11　5738 名苏士兰士兵的胸围分布频数比较

根据中心极限定理（Central Limit Theorem），只要是样本数很大，那就不管原来的总体的形式如何（只要它的平均数和标准差是有限的），样本的平均数就接近正态分布。

正态分布的公式是：

$$f\ (x)\ =\frac{1}{\sigma\ \sqrt{2\pi}}e^{-(1/2)(x-\mu)^2/\sigma^2} \tag{11.16}$$

μ 是平均值，σ 是标准差，f（x）是密度函数，也就是 y（纵线），即高度。按这个公式计算，

f（0）＝0.3938

f（1）＝0.2420

f（2）＝0.0540

但是我们通常更为关心的曲线下的面积，这需要用积分来计算其积累概率：

$$P_{(-\infty<X<\infty)}=\int_{-\infty}^{\infty}\frac{1}{\sigma\ \sqrt{2\pi}}e^{\frac{(x-\mu)^2}{2\sigma^2}}dX=1 \tag{11.17}$$

例如从 0 到 1 的面积应是 0.3413，从 0 到 2 的面积应是 0.4773……。从 0 到 $-\infty$ 的面积为 0.50，所以处于一个标准差位置的面积是 0.8413，处于两个标准差位置的面积是0.9773……。要计算积分比较麻烦，实际上我们只需查阅每一本统计学教科书的标准正态分布数值表，都可获得有关的面积的数值，见本书附录 A 的《正态曲线的面积和纵线表》。每一套数组都有其自身的平均值和标准差，这个表不可能为每一套数组提供专门的表，但是我们只要把数值转换成标准分（Z 分，见下一节），就可以查阅了。Z 分表示的是标准正态分布（平均值＝0，标准差＝1），例如 Z 分为 1，即表示一个标准差，查表可知从 0 到 1 的面积为 0.3413，较大部分面积为 0.8314，较小部分面积为 0.1587，纵线（y）为 0.2420。如果读者熟悉一些电算表（如 Excel），也可以在函数栏里找到 NORMDIST，就可计算密度函数（y）和积累概率，其指令为 NORMDIST（数值，数组的平均分，数组的标准差，TRUE 或 FALSE）。TRUE 所给的是积累概率，FALSE 所给的密度函数。如果已转换成 Z 分，则可输入 NORMSDIST（Z 分），得积累概率。

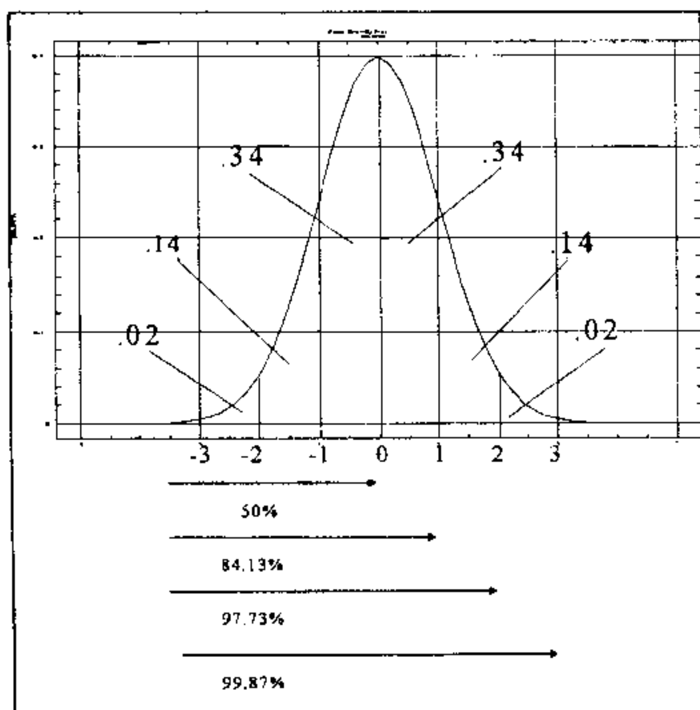


图 11.12 正态分布面积图

11.2.2.3. 标准分

标准分（Z 分）的计算公式是：

$$Z=\frac{X-\overline{X}}{s},\tag{11.18}$$

$\overline{X}$ 是一个数组的平均值，s 是它的标准差，所以我们可说标准分是根据平均数和标准差计算出来的一个转换分数。这是在实验和测试中使用甚广的一个概念，它用标准正态分布来表示分数所处的位置，使分数具有可比性。以我国高考为例，考生 A 在当年的某一科目考试里得到 75 分，而另一个考生 B 再去年同一科目的考试里得到 80 分。我们能否说考生 B 的水平要比 A 的高？或者说，我们能否说去年的 80 分高于今年的 75 分？有点测试常识的人都知道这个问题并非那么简单：因为不同的试题，或不同时间的平行试题的难度是不一样的，去年的 80 分是否高于今年的 75 分，要看去年试题的难度和今年试题的难度是否一样？如果难度一样，我们可以说 80 分高于 75 分。如果去年的试题比今年的容易得多，那么 80 分就不见得比 75 分高。换句话说，由于试题难度的影响，分数的含金量是不一样的。

这个问题的复杂性还在于考试的难度以及和它相联系分数是受两个因素所决定的，一个是试题本身的难易，另一个是考生的总体水平的高低。在这里我们先假定考生的总体水平没有明显变化（在大规模的考试里，人数众多的考生的水平不大可能在短期内发生重大变化），而去年考试的平均数是 78 分，而今年的考试的平均数是 70 分。去年拿 80 分的 B 比平均分高出了 2 分，而今年拿 75 分的 A 考生比平均分高出 5 分。我们就能形成一个概念，拿 80 分的考生不见得就比拿 75 分的考生水平高。但是这仅是一个直观的感觉，究竟这两个分数的差别有多大？我们就需要把它们转换成标准分（Z）：

假定这两年的考试的标准差都是 15，那么

$$Z_A=(75-70)/15=0.333$$
$$Z_B=(80-78)/15=0.133$$

标准分是和百分位相连的，如果标准分是 1，就意味着它处于比平均数高于一个标准差的位置，这个位置刚好比 84.13%的考生的成绩要高，而比它高的有 15.87% 。Z_A 的 0.333 的位置为 63%，而 Z_B 的 0.133 的位置则为 55%，可见 A 的实际水平比 B 的高，虽然他只有 75 分，而 B 却有 80 分。这里应该重复上面所强调的，这是假定这两年的考生水平稳定不变，平均数的变化反映了试题的难易程度的变化。在这个具体的例子里，我们假定两次考试的标准差是一样的，但是实际的情况并非如此。如果这两年考试的标准差不同，今年考试的离散性较小，标准差为 10，而去年的标准差为 20。A 和 B 的标准分就会有所变化：

$$Z_A=(75-70)/10=0.5$$
$$Z_B=(80-78)/20=0.1$$

Z_A 的 0.5 的位置为 69%，而 Z_B 的 0.1 的位置为 54% 。这是因为今年考试的分布不广，分数的权重要大些，比平均数多拿 5 分的权重应该大一些，而去年的分布较广，分数的权重没有那么大，故比平均数多拿 2 分的权重应该小一些。

标准分比原始分要科学得多，下例是 Guilford 提供的两个学生 A 和 B 的考试成绩的原始分和标准分的比较：

表 11.11　　A 和 B 的原始分和标准分比较

(1)	(2)	(3)	(4)		(5)	
			原始分		标准分	
测试	平均分	标准差	A	B	A	B
英语	155.7	26.4	195	162	1.49	0.24
阅读	33.7	8.2	20	54	−1.67	2.48
常识	54.5	9.3	39	72	−1.67	1.88
学能	87.1	25.8	139	84	2.01	−0.12
心理	24.8	6.8	41	25	2.38	0.03
总计			434	397	2.54	4.51
平均分			86.8	79.4	0.51	0.90

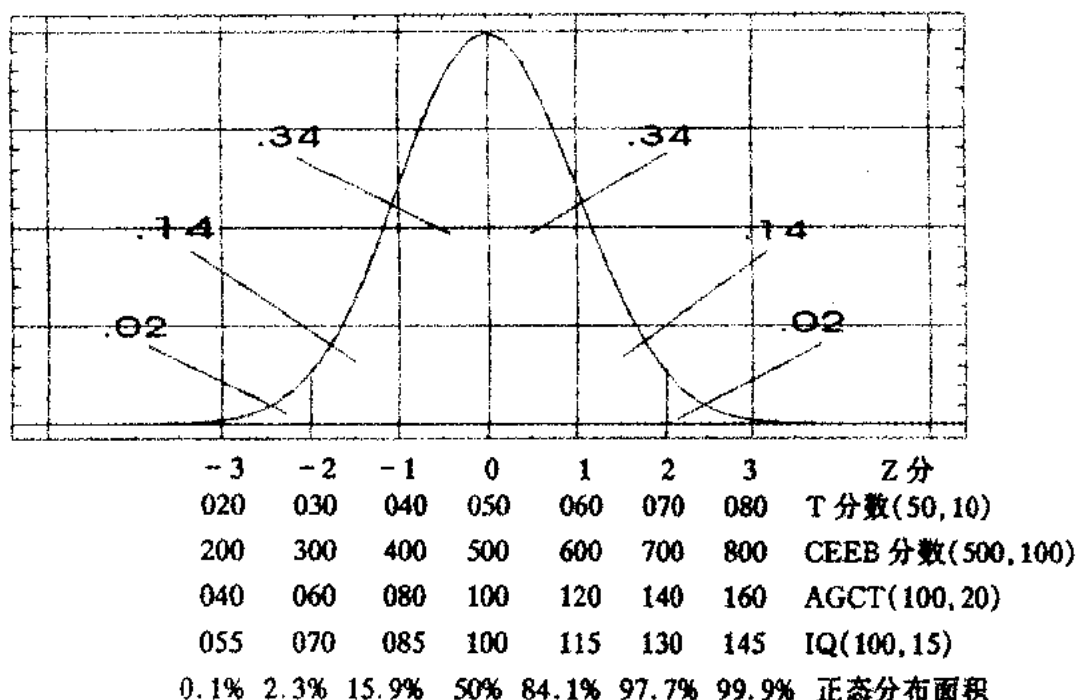
按原始分计算，A 比 B 的平均分高，但按标准分计算，则 B 的平均分高。这是因为标准分考虑了每科的平均分和标准差，作了权重处理。

使用标准分不大容易为人了解，而且操作情况起来不方便、因为它的平均分为 0，而且还

有正负值和小数点。于是教育学家想出不同的标准分常模，例如 Δ 量表：

$$\Delta = 13 + 4Z \tag{11.18}$$

这实际上是一个以 13 为平均分，以每个标准差为 4 的量表以代替了原来的以 0 为平均分，以 1 为标准差的量表。所以 Z 分为 0.333 就可转换为 14.332，Z 分为 0.133 就可转换为 13.532。这并没有把小数点去掉。另一个较流行的量表是以 500 分为平均分，以 100 分为标准差，所以 0.333 就可以转换为 500＋100×0.333＝533.3，而 0.133 就转换成 513.3。图 11.13 表示几种不同的标准分常模，它们的分数模式很不一样，但都可以转换成 Z 分，所表示的正态分布面积都是一致的。



−3	−2	−1	0	1	2	3	Z 分
020	030	040	050	060	070	080	T 分数(50,10)
200	300	400	500	600	700	800	CEEB 分数(500,100)
040	060	080	100	120	140	160	AGCT(100,20)
055	070	085	100	115	130	145	IQ(100,15)
0.1%	2.3%	15.9%	50%	84.1%	97.7%	99.9%	正态分布面积

图 11.13 正态分布和标准分

11.2.3. 概率分布的计算机运算

概率分布是统计理论的基础，在具体的运算方面并不复杂，所以有些软件包（如 SPSS）并没有这个功能。有的软件包则有概率分布的显示功能和拟合功能，如 STATISTICA。在 STATISTICA 里，我们可在主菜单 Module Switcher 里用 Switch to 去 Basic Statistics/Tables，再选择 Probability distribution calculator，便可有各种概率分布的图形显示，具有很强的功能，可任意输入一个值或概率，观察其分布，甚至可用上下的键来观察其分布的变化。我们也可以去 Nonparametrics/Distrib，再选择 distribution fitting，然后在表格上输入一组数据，指定一种分布，程序就能作出频数表，说明数据和这种分布的拟合程度，还可以作图，图 11.2 就是用这个软件作出来的。另一个软件包 Statgraphics 也有相同的功能。

11.3. 推断统计方法

11.3.1. 参数估计

我们在以上的章节里多次提到了样本和总体，推断统计方法就是要根据样本的参数来估计总体。为了把总体的平均数和样本的平均数区别开来，我们用 μ 来代表总体的平均数，σ 来代表总体的标准差，而用 $\bar{X}$ 来代表样本的平均数，用 s 来代表样本的标准差。

11.3.1.1. 总体参数的点估计

一般来说，我们对总体的一些参数是无法得悉的，例如中国人的平均身高是多少，我们不可能把十几亿的中国人的身高都来测量，而只能抽一些样本。问题在于怎样根据样本的统计量来估计总体参数。我们可以根据直觉说，样本的平均数 $\bar{X}$ 和总体的 μ 是相同的，但是我们的直觉有多准确？这两个平均数有多一致？$\bar{X}$ 有两个数学特性使我们用之于估量总体的 μ：首先，它是无偏的。在有些样本里，$\bar{X}$ 会小于 μ；在另外一些样本里，$\bar{X}$ 会大于 μ。但是把这些 $\bar{X}$ 加起来平均，这些偏差就会得到纠正。换句话说，如果我们把无数的样本的 $\bar{X}$ 加起来平均，样本的 $\bar{X}$ 和总体的 μ 就是一样的。这正是我们所希望的。其次，$\bar{X}$ 是一致的。这个意思是，如果估计建筑较大的样本的基础上，$\bar{X}$ 就会更接近参数的真值。事实上，我们经常都用样本的 $\bar{X}$ 来估计总体的 μ。从一个样本的数据中计算出一个单一的数值来估计总体参数就叫做点估计 (point estimator)。

11.3.1.2. 总体参数的区间估计

但是这个单一的数值有其自身的局限性，因为这个数值取决于一个单一的样本。另一个样本又会产生一个不同的数值。所以我们必须找出一种能够解释样本差异性的估计办法，这就是区间估计 (interval estimation)，它使用到置信区间 (confidence interval) 的概念。假定我们从一个包括了 100 个被试的样本中收集到一些反应时间数据，其平均数为 14.88 毫秒，标准差是 5 毫秒。又假定我们对总体的 μ 和 σ 并不知道，只知道样本数较大，$\bar{X}$ 是各种可能的样本平均数中一个随机观察量，而这些样本平均数都是属于一个正态分布总体的。在这种情况下，我们可以设想样本的 $\bar{X}$ 就是总体的 μ，但是必须设定一个误差的范围：

$$\mu=\bar{X}\pm 样本误差$$

这个样本误差也就是样本的标准差，表示为 $\sigma/\sqrt{n}$。在讨论正态分布时，我们谈到两个标准差 (Z 分) 的面积是 97.7%，假定我们取 1.96 个标准差，其面积就刚好是 95%。所以我们可以按照正态分布的原理来推定，这个总体的 95% 的样本在 $\mu\pm\sigma/\sqrt{n}$ 的区间将会有 $\bar{X}$（即在真正平均值的 1.96 个标准差之内）。在我们所举的例子里，原来的标准差为 5 毫秒，$\sigma/\sqrt{100}$ 就是 $5/\sqrt{100}=0.5$。原来的平均值为 14.88，因此其区间就是 $14.88\pm1.96\times0.5$，即总体的

真正平均值在于 13.9 到 15.86 之间。这就是置信区间，可表示为：

$$\overline{X}\pm Z\ (\sigma/\sqrt{n}) \tag{11.20}$$

Z 表示标准正态分布的标准差，若 Z=1.96，置信水平就是 95%。若 Z=2.58，置信水平就是 99%。

置信区间和样本数有很大的关系，样本有 10 个被试或 1000 个被试，其置信区间会很不一样。

	n=10	n=1000
样本标准差	1.58	0.158
置信区间	±3.1	±0.31
总体 μ	11.78～17.98	14.57～15.19

可见样本数越大，区间就越小；样本数越小，区间就越大。这个样本平均值的样本标准差又可称为样本平均值的标准误 (standard error)。上面说过，按照中心极限定理，样本必须够大，其平均值才能正态分布，那么怎样样本够大呢？如果我们所考察的变量本身是正态的，那么不管样本数有多大，它都会是正态的。如果变量本身不是正态的，但是它的总体的直方图只有一个众数，而且基本上是对称的，那么 20 个左右的样本数就能保证样本平均值是正态的。只有变量是非常偏态的，而且直方图是明显双峰的，才要求的大样本，但 100 个样本数是足够的了。

11.3.1.3. 比例的估计

有时我们所获得的数据以比例的形式出现，比例其实也是样本的平均数，而且如果样本数够大，它也是正态分布的。为了了解比例的置信区间，我们必须知道样本的标准差，然后按 $\sigma/\sqrt{n}$ 来计算。但是有一个更直接的办法：

$$p\pm\left\{1.96\sqrt{\frac{p\ (1-p)}{n}}\right\} \tag{11.21}$$

例如有人调查了英国儿童语言发展到某一个阶段会使用一些过度概括的过去时态，如说 bringed，runned 等。结果是正确使用不规则的过去时态的频数为 987，用过度概括的过去时态的频数为 372。两者合共为 1359，而使用后者的比例为 372/1359=0.274。所以 95%的置信区间就是 $0.274\pm1.96\times\sqrt{(0.274\times0.736)\ /1359}=0.274\pm0.024$，即真正的平均数在 0.25 到 0.298 之间。但是这个置信区间所给的范围偏窄，所以有的统计学家主张作一些修正 (见 Woods，1986)，其公式如下：

$$p\pm\left\{1.96\sqrt{\frac{p\ (1-p)}{n}}+\frac{1}{2n}\right\} \tag{11.22}$$

在我们的这个例子里，差异不很大，主要是样本数很大。

11.4. 假设检验

11.4.1. 使用置信区间来检验假设

在上一节里，我们介绍了使用置信区间来估计总体参数，这个置信区间也可用来评估参数的假设数值。假定我们对我国学习英语的学生的词汇量感兴趣，并且开展了调查，了解到英语专业的学生基础阶段结束后掌握的词汇量为 6000（桂诗春，1988），而我们在某一个学校对 100 个同类的学生用同一个方法调查，结果是 6200，标准差为 1000。我们能否根据这个学校的样本来假设这个总体和我们调查的平均词汇量是一样的？当然这个学校的平均词汇量比我们调查的词汇量的大，但是我们调查中也有一些样本的词汇量是超过 6000 的。换句话说，虽然这个学校的样本的词汇量为 6200，它会不会在我们调查的词汇量的置信区间里面？也就是说，这两个数值的差别是否仅是样本的差别？如果我们抽另外一组学生来观察，其平均词汇量会不会更接近 6000？

在这里我们只知道总体的 μ，而不知道总体的 σ；但我们知道样本的标准差 s，因为样本数较大，我们可认为 s 是最接近 σ 的，是无偏的统计量。公式（11.20）可写为：

$$Z=\frac{\overline{X}-\mu}{s/\sqrt{n}} \quad (11.21$$

根据（11.20）可算出：

$$6200\pm 1.96\times 1000/\sqrt{100}=6200\pm 196\ (6004,\ 6296)$$

我们所选定的置信度为 95%，这就是说，在 95% 的情况下，真正的平均值在 6004 到 6296 之间，现在我们所假设的真正平均值为 6000，所以样本值 6200 只有 5% 的可能性是真正平均值。这就是说，如果我们抽 20 个样本，有 19 个样本的平均值都会在 6004 到 6296 之间，只有 1 个样本不在这个区间。因此我们可以有理由认为这个学校的样本和我们所调查的单位并非属于同一总体。

11.4.2. Z 检验

从提出假设的角度看，视乎置信区间有没有包括 6000，可以有不同的结论：

表 11.12　假设检验的四种可能结果

样本结果	真正的情况	
	总体平均值为 6000	总体平均值并非 6000
区间包括 6000	正确决策	第二类错误
区间不包括 6000	第一类错误	正确决策

这就是说，如果总体平均值为 6000，而区间包括 6000，我们可下结论：数据和我们的假设相符，即样本属于总体。这也就是说，无差别假设属真，而我们又接受无差别假设，因此这是正确的决策。如果区间不包括 6000，这就等于说，我们怀疑无差别假设是否属真，通常的说法就是，我们认为假设是假的而加以拒绝，就会犯第一类错误（亦称 α 错误）。第一类错误是“无差别假设属真而我们又加以拒绝的错误”。这就是说，20 个样本中有 19 个样本的区间都不会包括 6000，只有 1 个（其概率为 5%）有可能，而我们却拒绝了 19 个样本的区间，反而接受那唯一的一个的区间。

如果总体的平均值并非 6000，置信区间自然也不应包括 6000，这就等于说无差别假设属假，而我们又加以拒绝。这是正确的决策。但如果我们还是把 6000 放在置信区间，即接受了无差别假设，那就犯了第二类错误（亦称 b 错误）。第二类错误是“无差别假设属假而我们又加以接受的错误”。一般来说，犯这类错误的概率是不知道的，它取决于真正的平均值、样本数、总体的方差。

因为假设可真可假，所以就有四种结果（见表 11.12）：两种结果使我们正确估计情况，另外两种结果导致错误的结论。表 11.12 可以概括为：

	无差别假设属真	无差别假设属假
接受无差别假设	正确决策	第二类（β）错误
拒绝无差别假设	第一类（α）错误	正确决策

我们可再看公式（11.21），这个公式的意思是：找出 $\bar{X}$ 和 μ 的绝对差，再除以标准误，如果 Z 小于我们所用来计算置信区间的 Z（如 1.96）就可以说区间包括了 μ，否则就不能包括。所以，

$$(6200-6000)\ /1000/\sqrt{100}=2$$

2>1.96，故 6000 不包括在 95%的区间里。假定我们选定的置信水平为 99%，其 Z 为 2.58，2<2.58，故可把 6000 包括在区间里面。若用（11.20）公式计算，99%的置信区间为 6200±58，即在 5942 到 6458 之间，而 6000 包括在置信区间之内。我们可以把这种情况表示为 $0.01<P<0.05$，P 可以理解为犯第一类错误的概率，也可以称为显著性水平。用统计学的术语来说，$\mu=6000$ 可以在 5%显著性水平上加以拒绝，但不能在 1%显著性水平上加以拒绝。

“在 5%显著性水平上”实际上是说，根据样本推定所提出的平均值并不包括在 95%置信区间。公式（11.21）所表示的 Z 值可以叫做检验统计量（test statistic）。每一个检验统计量都是一个随机变量，因为这个值取决于一个随机抽样程序的结果。如果我们重复使用这个程序，每次抽不同的样本，最后我们便能取得检验统计量数值的一个随机样本，并根据它的直方图作成一条曲线。选择检验统计量的前提是：在检验的假设是真的情况下，它的直方图的数学模型是已知的。例如在我们的词汇量调查里，我们已知总体的平均值是符合标准正态分布。所以当我们选定 Z=1.96 时，$\mu=6000$（或在 6004 到 6296 之间）的概率为 95%，而大于或小于 μ 的概率各为 2.5%。

在检验假设[1]时，我们往往先提出一个无差别假设，表示为H_0。然后我们再去检验H_0：$\mu=6000$，而我们已知中国的英语学生的词汇量的掌握是呈正态分布的。在上述的计算中，我们得知Z＝2，如果Z＞1.96时，我们就应拒绝H_0，而接受备择假设H_1。备择假设可有三种：

(i) H1：m＜6000
(ii) H1：m＞6000
(iii) H1：$\mu\neq6000$

我们不妨再看公式（11.21）中的样本平均值减μ，既然样本平均值可以大于或小于μ，所以最后得到的Z也可能是正值或负值。如果(i)是真的，总体的平均值应小于6000分，这个总体的样本平均值也会小于6000。在这种情况下，Z就会是负值。因此大的负Z值支持H_1，而不是H_0。如果Z值不是负值，哪怕它再大，也不支持这个备择假设。如果(ii)是真的，情况就刚好相反，大的正Z值也支持H_1，而不是H_0。如果(iii)是真的，Z可以是正的，也可以是负的，视μ而定，只要是Z够大，我们都可以拒绝H_0。

表 11.13

单侧检验	双侧检验
0.05 临界值＝1.65	0.05 临界值＝1.96 (0.025)
0.01 临界值＝2.33	0.01 临界值＝2.58 (0.005)

在这种情况下，我们往往需要查表来决定哪些检验统计量的数值足以支持H1，这些数值可称为检验统计量的临界值。这些表通常都告诉我们检验的显著性水平。在查表时，我们应注意单侧或是双侧检验。当备择假设并不指明样本平均值和总体平均值的差的方向，如(iii)的$\mu\neq6000$时，我们需要使用双侧检验。在这种情况下，我们只关心有没有区别，而并不关心这个区别是正的还是负的。但是有时我们想了解某一方向的差别，如样本平均值是否大于（或小于）总体的平均值，就可以用单侧检验。

双侧和单侧所要求的Z值是不同的，可参看表11.13。

11.4.3. t 检验

我们在上面谈到中央极限定理和Z检验，指的是大样本的平均值呈正态分布。但是如何根据小样本来估计总体呢？

假定某出国人员培训部作强化英语训练，根据过去的经验，入学水平为70分的学员经过半年培训后可达78分。但是有一个10个人的小班只培训了3个月，便需要出国，结业前的考试结果表明，他们的平均值为72.8，标准差为14.6，我们想知道他们是否达到训练所要求的水平？在这种情况下，无差别假设（H_0）是m＝78，而备择假设（H_1）则是m＜78。使用公式（11.21）计算，检验统计量为－1.13。但是这个例子和上面的词汇量调查的例子不同，样本数只有10个，我们很难说这是符合正态分布的。

统计学家Gossett于是提出一种概率分布，叫做Student分布，Student是他发表研究文

---

[1] 假设的检验还可参看10.2.6.。

章时所用的笔名。现在大家简单地称为 t 分布。t 分布的形状和正态分布相同，但是在 n 较小时，分布得较为扁平。然后随着 n 的数目的增加，就往正态分布不断靠拢。这就是说，用 t 检验来处理大样本时，t 的临界值和 z 的临界值差不多是相同的。因为它们很相近，所以一些统计人员不管样本大小，都用 t 检验。

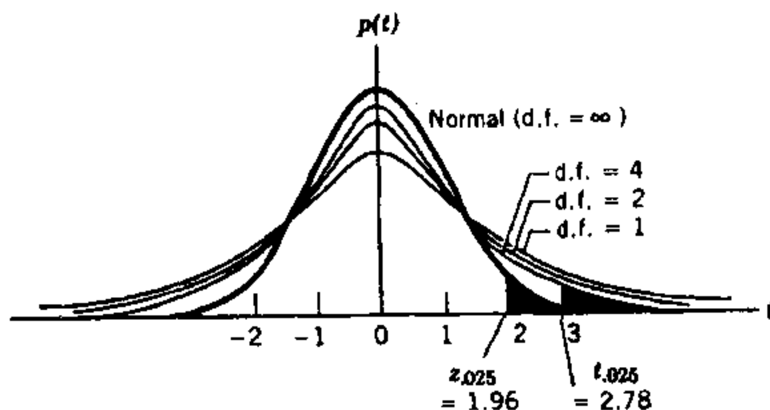


图 11.14　自由度为 1，2，4，∞时 t 分布的图形

因为 t 分布和样本数量有关，这就需要增加一个自由度（degree of freedom）的概念，自由度表示为 n－1。t 分布的样本函数是：

$$t=\frac{\overline{x}-\mu}{s/\sqrt{n}} \tag{11.22}$$

这个样本函数服从自由度（n－1）的 t 分布，换句话说，不同的自由度有不同的 t 分布。我们用 n－1，而不是 n 的理由是：我们的样本数小，总体数大，而我们使用这个统计量来估计总体参数。样本的分数的离散程度要大于总体分数的离散程度，用 n 来除样本的数值不能很好地估算总体的参数，于是统计学家们使用 n－1。下面是一个 t 分布的部分数值表：

表 11.14　　t 概率分布表（部分）

自由度 (n－1)	单侧概率					
	0.10	0.05	0.025	0.01	0.005	0.0005
	双侧概率					
	0.20	0.10	0.05	0.02	0.10	0.001
1	3.08	6.31	12.7	31.8	63.7	637
8	1.40	1.86	2.31	2.90	3.36	5.04
9	1.38	1.83	2.26	2.82	3.25	4.78
10	1.37	1.81	2.23	2.76	3.17	4.59
20	1.32	1.72	2.09	2.53	2.85	3.85
60	1.30	1.67	2.00	2.39	2.66	3.46
∞	1.28	1.64	1.96	2.33	2.58	3.29

如果我们把这个表的无穷大（∞）那一行和表 11.13 的临界值相比，就可看到当 t 分布无穷大时，它就是正态分布的；但是当自由度较小时，临界值就有所不同。单侧和双侧概率又有所不同。在上面的例子里，培训了 3 个月的学生得到平均值 72.8，t 值和统计检验量是一样的，−1.13。负值表示样本的平均值低于总体的平均值，我们只想了解样本的平均值虽然低于总体的，但究竟是否有差别，这是单侧检验，查表 11.14 自由度为 9 的那一行可见，必须>1.83（或<−1.83，一般可以不考虑 t 值的正或负，因为这只表示差别的方向）才能拒绝无差别假设。现在 1.13<1.83，我们不能接受备择假设 $H_1: \mu<78$。所以我们可认为这 10 个人的平均值和总体的平均值在 0.05 的显著性水平上并没有差别。但是这里应该明确两个概念：(1) 这只是说明，从检验假设的角度看，无差别假设有 95%的可能性是真的，平均值为 72.8 的样本和总体的平均值 78 分只有 5%的机会是有差别的。这并非说 72.8 分等于 78 分。(2) 这只是平均值的比较，并非说这 10 个人每个人都达到要求。

t 检验是比较平均值的最常用的一种检验方法，检验所得到 t 值的显著意义通常都可以查表，如附录 B 的 t 概率分布表，很多统计软件在给出 t 值后，往往也提供其概率。

上面的例子可以说是单样本的 t 检验。其实平均值的比较还有不同的情况。

11.4.3.1. 配对 t 检验

配对 t 检验（Paired t-test）比较一个样本的不同平均值，也可以称为重复检验。我们可以比较一个样本实验前后的平均值有何差别，也可以比较一批学生在两个不同的测试中的成绩有何差别，也可以比较不同的阅卷员对一批作文题的评分的平均值有何差别。例如：

表 12.15 5 个学生参加两次考试的平均值比较

	观察值		分数差
学生	X_1	X_2	$X_D=X_1-X_2$
A	60	63	−3
B	65	67	−2
C	50	70	−20
D	70	70	0
E	80	82	−2
平均值	65	70.4	−5.4
标准差 (s_D)			8.234076
t 值			−1.46644

这是 5 个学生参加一个教学法实验前后两次考试成绩的平均值的比较，第一次的平均值为 65 分，第二次为 70.4。表面看来成绩确有变化，我们能否推断说学生的进步是参加教学法实验的结果？查 t 分布表（见附录 B）知道自由度为 4，显著性水平为 0.05 时的临界值为 2.132，现在 t 值为 1.46<2.132，我们不能拒绝无差别假设。这就是说，70.4 分虽然比原来的 65 分高了 5.4 分，但是样本数太少，产生误差的情况太多(例如再找另外 5 个学生做实验，说不定就不会增加 5.4 分)，我们如果拒绝无差别假设，就会犯了第一类错误：不能下结论时仓促地下了结论。计算这种配对 t 检验的公式是：

$$t=\frac{\bar{x}_D-\mu}{S_D/\sqrt{n}} \qquad (11.23)$$

$\bar{x}_D$ 为两组平均值之差，s_D 为其标准差。配对 t 检验的置信区间可用下列公式计算：

$$\Delta=\bar{x}_D \pm t_{0.025}\frac{S_D}{\sqrt{n}} \tag{11.24}$$

在这个例子里，$\Delta=-5.4\pm2.78\times8.234/\sqrt{5}=-5.4\pm10.235$，我们可写成 CI（-15.635，4.835）。这就是说第一组的平均值 65 分可以比第二组的平均值 70.4 分低 15.635 分或高 4.835 分，置信区间之所以大是因为样本数太少。$t_{0.025}$（2.78）是设定置信度为 95%后查阅 t 值分布表（附录 B）所得的临界值。

11.4.3.2. 独立样本的 t 检验

独立样本 t 的检验和配对的 t 检验不同，它牵涉到两个独立的样本的平均值的比较。配对 t 的检验是对一个样本参加两次测试的平均值做比较；而独立样本 t 检验是对两个样本参加一次测试的两个平均值做比较。因此在自由度方面就会不一样，在上面的例子里，5 个学生参加两次测试，其自由度为 4。如果我们对两个样本（各为 5 名学生）的两个平均值做比较，其自由度是 $n_1-1+n_2-1=8$。和配对 t 检验不同的另一点是这两个样本的数目不一定相等。例如两个小班的学生经过两年的基础阶段的教学后，参加一个水平考试。结果是一个班的平均值为 81 分，另一个的平均值为 77 分，我们能否推断说这个班的水平比另一个班的高？这两个班是独立样本，要用独立样本 t 的检验。

这种检验方法的公式略有不同：

$$t=\frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{(n_1-1)\ s_1^2+\ (n_2-1)\ s_2^2}{n_1+n_2-2}\ (\frac{1}{n_1}+\frac{1}{n_2})}} \tag{11.25}$$

表 11.16　　两个样本的平均值比较

分数	甲班	乙班
	65	70
	67	78
	65	76
	74	65
	75	84
	80	85
	82	78
	85	90
	87	92
	90	92
平均值	77	81
标准差	9.237604	9.237604
t 值		-0.96825

用这个公式计算出来 t 的值为-0.968，自由度为 18，查附录 B 表可见 0.05 显著水平的临界值为 2.101，现 0.968<2.101，因此我们接受无差别假设，即考虑到一些随机误差，不能认为乙班的学生的水平就一定高于甲班。

独立样本 t 检验的置信区间可用下列公式计算：

$$(\mu_1-\mu_2)\ =\ (\bar{x}_1-\bar{x}_2)\ \pm t_{0.025}s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \tag{11.26}$$

计算结果是 CI（-13.84，5.84）。

t 检验是一种相当稳健的检验，因为我们可以不必考虑平均值是否正态分布。但是 t 检验也需要有一些假定：(1) 被试在两个对比组的安排是随机的；(2) 独立变量的分数是连续性的，

而且变量只有两个层面（平均值）；(3) 总体的分数的方差是相等的，而且分数是正态分布的；(4) 我们不能对几个组的平均值做交叉比较，例如比较平均值 1 和平均值 2，比较平均值 1 和平均值 3，比较平均值 2 和平均值 3。这样做会误导我们拒绝无差别假设，因为交叉的比较会使概率提高。假如我们设定的显著性水平为 α，则 $\alpha=1-(1-\alpha)^c$，c 是交叉比较的次数。所以比较了 4 次，$\alpha=1-(1-0.05)^4=0.18$。我们所定的显著性水平就不再是 0.05 而是 0.18 了。要做交叉比较应该使用后面要谈到的方差分析（ANOVA）。

11.4.3.3. t 检验的计算机运算

SPSS 和 STATISTICA 作 t 检验的做法基本相同。在 SPSS 里，要从主菜单到 Compare means，然后再选择 Paired Samples T Test（配对 t 检验），还是 Independent Samples T Test（独立样本的 t 检验），如表 11.15 的数据可按两列输入 SPSS 的工作表，运算后的结果如图 11.15。

表上所给的 corr 是两组数据的相关系数。独立样本 t 检验的数据输入办法和配对 t 检验的办法不一样，因为两组的样本数可以是一样的，也可以是不一样的，故我们必须有一个变量说明数组，如：

1	65
1	67
⋮	⋮
⋮	⋮
1	90
2	70
2	78
⋮	⋮
⋮	⋮
2	92

```
05 Jul 96 SPSS for MS WINDOWS Release 6.0                    Page 1
                 ------ t-tests for paired samples ------
           Number of           2-tail
Variable     pairs     Corr     Sig      Mean       SD       SE of Mean
---------------------------------------------------------------------------
VAR00003                                 65.0000    11.180     5.000
               5      0.678   0.209
VAR00004                                 70.4000     7.092     3.172
---------------------------------------------------------------------------
        Paired Differences
Mean         SD        SE of Mean       t-value    df   2-tail Sig
--------------------------------------------------------------------
-5.4000      8.234     3.682            -1.47      4     0.216
95% CI (-15.628, 4.828)
05 Jul 96 SPSS for MS WINDOWS Release 6.0                    Page 2
```

图 11.15

```
08 Jul 96 SPSS for MS WINDOWS Release 6.0                    Page 1
t-tests for independent samples of VAR00003
                 Number
  Variable      of Cases      Mean       SD       SE of Mean
------------------------------------------------------------------
VAR00001
VAR00003 1     10     77.0000     9.238     2.921
VAR00003 2     10     81.0000     9.238     2.921
------------------------------------------------------------------
        Mean Difference = -4.0000
        Levene's Test for Equality of Variances: F=0.010   P=0.920
t-test for Equality of Means                                 95%
Variances   t-value    df    2-Tail Sig    SE of Diff     CI for Diff
------------------------------------------------------------
Equal       -0.97      18    0.346       4.131      (-12.681, 4.681)
Unequal     -0.97      18    0.346       4.131      (-12.681, 4.681)
------------------------------------------------------------
08 Jul 96 SPSS for MS WINDOWS Release 6.0                    Page 2
```

图 11.16

11.4.4. χ^2 检验

χ^2（念作卡方）检验是一种非参数检验[①]，它是在总体分布未知的情况下，根据样本的频数分布对总体分布所作的假设检验。χ^2 检验主要是检验频数的分布和某个概率分布模型是否一致，这叫做拟合优度（goodness of fit）。χ^2 检验统计量的主要形式为：

$$\chi^2=\sum\frac{(f_o-f_t)}{f_t} \tag{11.27}$$

f_o 为实际频数，f_t 为理论频数。χ^2实际上是实际频数和理论频数差的平方和理论频数之比。从（11.27）可看出，χ^2 有如下的特点：(1) χ^2 永远是一个正值，因为不管实际频数和理论频数值差是正还是负值，其平方永远是正的。(2) χ^2 分布曲线也和 t 分布曲线一样，是由无限曲线组成的曲线族，随着 n 的不同，χ^2 分布曲线也就不同，因此其样本函数也服从自由度（n－1）的 χ^2 分布。χ^2 的大小随实际频数与理论频数差的大小而变化。(3) χ^2 值具有可加性。

χ^2 的分布如图 11.17。

这个分布的特点是（1）它是左右不对称的，呈正偏态，左面较陡，右面较平，右侧无限延伸，但永不与基线相交。(2) 当 n<2 时，曲线是下降的，但当自由度不断增大时，分布就

① Nonparametric tests 用来处理名义变量和次序变量，它只需要频数。和 Z-检验和 y-检验不同，它不要求样本所属的总体呈正态分布。

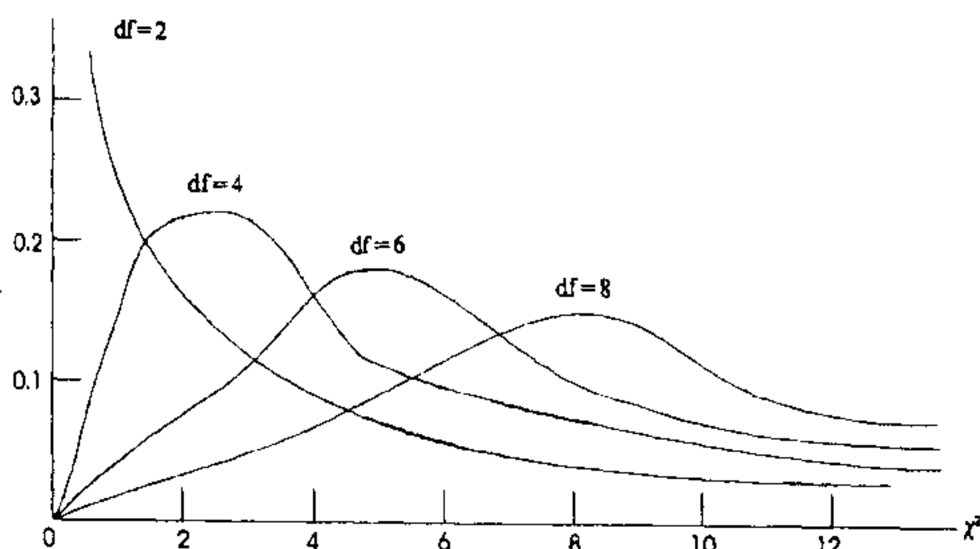


图 11.17 自由度为 2，4，6，8 时 χ^2 分布图

趋于对称。

χ^2 也可查表（附录 C），因其左右侧的概率是不同的，故列表的方式也不同：有的左右侧分开列表，有的从左到右列表，有的只给高低的百分位。一般来说，最后的一种就足够了，因为我们往往需要知道的不是 0.05 和 0.10，就是 0.95 和 0.99 的概率。下面是这样的表的部分数值：

表 11.17　　χ^2 分布的高低百分位数值（部分）

df	p							
	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
5	0.554	0.831	1.145	1.61	9.236	11.070	12.832	15.086
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
30	14.553	16.791	18.491	20.599	40.256	43.773	46.979	50.892
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154

11.4.4.1. 单向表的 χ^2 检验

单向表是把频数按一种分类标准编制成的表，例如按年龄整理为老年、中年、青年、少年；按受教育情况整理为研究生、大学、中学；按区间整理成不同的分数档等等。对于单向表的数据进行 χ^2 检验就是单向表的 χ^2 检验。

例如某大学对 120 应届毕业生的毕业后从事工作的志愿作调查，希望到贸易业部门的 60 人，希望到政府机关的 30 人，希望继续学习的 15 人，希望到教育部门的 15 人。我们想了解学生的工作志愿是否有明显的差别。

(1) 第一步是提出假设:

无差别假设为 H_0: $p_1=p_2=p_3=p_4$

备择假设为 H_1: p_1、p_2、p_3、p_4 中至少有一对是不相等的。

(2) 第二步是作出样本函数:

除了列出实际频数外,还必须算出理论(即期望)频数:现有120人,4种志愿,如果他们的志愿是无差别的,理论频数就是120/40—3。

表 11.18　　χ^2 值计算表志愿

志愿	f_o	f_t	f_o-f_t	$(f_o-f_t)^2$	$(f_o-f_t)^2/f_t$
贸易	60	30	30	900	30
政府部门	30	30	0	100	0
继续学习	15	30	−15	625	7.5
教育	15	30	−15	625	7.5
χ^2					45

(3) 决定自由度和显著性水平。自由度为 k 1=3;显著性水平选 $\alpha=0.05$。

(4) 比较样本函数所得的 χ^2 (45) 和 χ^2 分布表中的临界值 (7.815),45>7.815,可拒绝无差别假设,认为学生的志愿有显著性差别。这可以表示为:$\chi^2=45>\chi^2_{(3,0.05)}=7.815$。

单向表的 χ^2 检验可用来检验频数分布是否符合正态,例如表 11.4 给出各个区间的频数,我们想知道其分布是否正态。无差别假设为频数分布与正态分布无差别。这首先必须求出理论频数,方法是将分数转换成 Z 分,现平均值为 65.18,标准差为 16.23,故 Z=(x−65.18)/16.23。根据 Z 分便可查出在正态分布中的累积面积,然后算出每个区间的面积,这就是理论的概率;再用总数乘以理论概率,便是理论频数。有了实际频数和理论频数,便可算出 χ^2 值。

表 11.19　　检验表 11.4 数组是否属于正态分布

区间	频数	Z 分	区间面积	理论频数	$(f_o-f_t)^2/f_t$
21~30	3	−2.47566	0.006649	0.664947	8.199854
31~40	5	−1.85952	0.024827	2.482724	2.552309
41~50	13	1.24338	0.075388	7.538792	3.956176
51~60	16	−0.62723	0.158388	15.83884	0.00164
61~70	28	−0.01109	0.230323	23.03225	1.071476
71~80	19	0.605052	0.231852	23.18524	0.75549
81~90	8	1.221195	0.161566	16.15659	4.117825
91~100	8	1.837338	0.077926	7.792617	0.005519
	100			$\chi^2=$	20.66029

查附录表 C 可见自由度为 5,显著性水平为 0.05 的 χ^2 临界值为 11.1,$\chi^2=20.7>\chi^2_{(5,0.05)}=11.1$。由此可作判断,拒绝无效假设,这个频数分布不属正态分布。这里需要解释一下为

什么自由度为5，而不是7。自由度可以理解为我们有多少件独立的信息可作为检验假设的依据，我们在本例里有8个区间的频数要作比较，但是最后一个是全部频数减去前面7个的和，所以实际上只有7个。但是我们在估计总体的平均值和标准差时，又使用了两件独立的信息，这也需减去，故自由度为k－1－2＝5。自由度的概念不容易解释清楚，在以后的讨论中我们也不准备逐一解释，只说明哪一种统计量需要怎样计算自由度。

11.4.4.2. 双向表的χ^2检验

双向表把频数按两种分类标准进行整理和排列，例如我们把被试按老、中、青分类，再理解他们是否吸烟，或把学生按年级分类，再了解他们在英语学习中犯几种常见的语言错误的情况等等。双向表的χ^2检验是独立性模型的检验，其目的是了解两个因素是否互相独立的，如果不是互相独立的，就是互相有联系。在各种问卷调查中我们往往需要使用这种方法来决定被调查人的各项反应是否有显著性差异。例如我们曾在广州外国语学院调查英语专业和非英语专业学生对不同课程的喜爱程度（桂诗春，1986），其目的就是了解两种学生对所开设的课程的态度是否一致，即互有联系，这就是我们的无差别假设：

表11.20　　学生喜爱课程频数表

	外语基础课	外语选修课	汉语	政治	国际关系	第二外语	体育	其他	总计
英语专业	274	114	18	5	54	16	20	10	511
非英语专业	169	52	30	4	39	13	25	25	357
总计	443	166	48	9	93	29	45	35	868

我们必须首先按照下列公式计算出理论频数。

理论频数＝行总数×列总数/总数　　(11.28)

表11.21　　学生喜爱课程理论频数表

	外语基础课	外语选修课	汉语	政治	国际关系	第二外语	体育	其他	总计
英语专业	261	98	28	5	55	17	26	21	511
非英语专业	182	68	20	4	38	12	19	14	357
总计	443	166	48	9	93	29	45	35	868

用公式(11.25)计算结果，χ^2为34.4。自由度（df）＝(a－1)(b－1)，a是行数，b是列数，故df＝7。$\chi^2=34.4>\chi^2_{(7,0.05)}=14.1$，因此我们可拒绝无差别假设，认为两类学生对所开设的课程的态度是不同的。但是我们还不知道哪些组的差异是显著的，哪些组的差异是不显著的。要进一步了解这

表11.22　　列联表

	外语基础课	外语选修课	总计
英语专业	a＝274	b＝114	a＋b＝388
非英语专业	c＝169	d＝52	c＋d＝221
总计	a＋c＝443	b＋d＝166	N＝609

些差异，就必须作 2×2 的双向表（也可称为列联表[1]）χ^2 检验。

列联表的 χ^2 检验可以用一种简单的方法，无须计算理论频数，其结果和公式（11.26）是一样的，其公式是：

$$\chi^2=\frac{(ad-bc)^2 \cdot N}{(a+b)\ (a+c)\ (b+d)\ (c+d)} \tag{11.29}$$

结果是 2.43，自由度＝（2－1）（2－1）＝1，$\chi^2=2.43<\chi^2_{(1,0.05)}=3.84$，因此我们不能推翻无差别假设。可见两类学生虽然对各种课程的开设的态度是不一致的，但他们对外语基础课和外语选修课的态度却是一致的。

11.4.4.3. χ^2 检验要注意的问题

χ^2 检验使用得很广泛，但有时容易忽略一些应该注意的问题，导致检验失效：

1. 理论频数太少。在 χ^2 检验里，理论频数不能太少，通常应该多于 5。要解决这个问题，有两种办法：一种办法是看看变量是否分得太细；如果太细，就可以把一些范畴合并。当然更彻底的办法是想法收集更多数据。另一种办法是继续进行 χ^2 检验，但是要知道这会产生一个当无差别假设为真时大于正常的统计量，换句话说，会增加犯第一类错误的可能性，所以我们下结论时应该小心谨慎。Yate 的校正公式是：

$$\chi^2=\sum\frac{(|f_o-f_t|-0.5)^2}{f_t} \tag{11.30}$$

2. 观察的独立性。在进行双向表 c2 检验时，我们强调独立性模型，这就是说，我们所分析的数据必须是互相独立的。Woods（1986）等举了一个例子来说明数据必须互相独立的重要性：一个研究者观察两组英语学习者在特定语境中掌握英语复数的语素（/s/，/z/，/iz/）的情况，一组是在课堂上正式学习英语的，而另一组则不是。然后研究者访问每个被试 1 小时，把他们正确使用语素和没有使用的情况记录下来，如表 11.23：

表 11.23　　两组学生掌握英语复数的语素的调查

	正式学习组		非正式学习组	
	使用语素	没有使用语素	使用语素	没有使用语素
	32	28	42	36
	112	24	39	28
	106	39	31	29
	42	40	55	37
	26	24	62	55
组总计	318	155	229	185

[1] Contingency Table，亦称为相依表、四格表。

我们不能使用组总计来进行 χ^2 检验，首先是有一个组的两个成员的使用频率特别高，会影响组的总分，更重要的是用组的总分来计算 χ^2 检验是错误的，因为每个人在一个小时内每次使用这个语素的行为并非独立于其他时候的行为。

3. 在一项研究中作几种交叉检验的问题。例如有一项研究观察 60 个学生的 6 种语法错误，它把每一种错误都列表来比较两个组在 0.05 显著水平上的差异，这等于交叉检验，结果显著水平就再不能是 0.05，而是 $1-(0.95)^6=0.26$（见 12.4.3.2）。其次是当我们观察同一文本里的不同语言变量时，这些变量可能不是独立的。

4. 使用百分比。有时频数不好比较，我们往往把频数转变为百分比，但是要注意用百分比作 χ^2 检验的结果并不准确，χ^2 检验必须用原始的频数。如果需要用百分比来比较，也必须先用频数来计算 χ^2，然后再把频数转换称百分比。

11.4.4.4. χ^2 检验的计算机运算

行	列	观察频数
1	1	274
1	2	114
1	3	18
1	4	3
1	5	54
1	6	16
1	7	20
1	8	10
2	1	169
2	2	52
2	3	30
2	4	1
2	5	39
2	6	13
2	7	25
2	8	25

Var1	Var2
1	60
2	30
3	15
4	15

单向表 χ^2 检验比较简单，只需要输入观察数值和期望数值，在 EXCEL5 或 STATISTICA 上都可以计算。在 SPSS，则需要输入两个变量，第一个是范畴（caterogy），是数值的编号，第二个是观察频数。然后在主菜单里去 Data/Weight Cases，把观察频数定义为频数，再去 nonparametrics/Chi-Square，进入菜单后，在 Test Variable List 里输入范畴变量的名称，再按 OK，就可得到 χ^2 值。例如表 11.18，应按右表的方式输入。

如果是双向表的 χ^2 检验，则要按交叉列表的方式输入数据，再去 Statistics / summarize/ crosstabs，进入 crosstabs 后，再按 Statistics 的按键，选 Chi-Square，例如表 11.20 的数据应按左表的方式输入。输入数据后，照样要到 Data / Weight Cases 那里去把观察频数加权，然后再计算，其结果如下：

```
08 Jul 96 SPSS for MS WINDOWS Release 6.0
        [略去交叉列表]
        Chi-Square                    Value        DF        Significance
Pearson                               34.63639     7         .00001
Likelihood Ratio                      34.46770     7         .00001
Mantel-Haenszel test for              16.64920     1         .00005
linear association
Minimum Expected Frequency —   3.702
Cells with Expected Frequency<5—    1 OF    16 (6.3%)
Number of Missing Observations: 0
08 Jul 96 SPSS for MS WINDOWS Release 6.0
```

图 11.18

11.5. 相关系数与线性回归

在上面的章节里，我们讨论了变量的差异（如Z检验和t检验）以及独立性模型的检验（如χ^2检验），但是变量之间还有另一种关系，那就是依存关系（interdependence）。我们想了解的是一个变量是怎样和别的变量联系的，统计学家把这叫做回归（regression）。例如吸烟和癌病有些什么关系，肥料和作物产量有些什么关系，母语和外语学习有些什么关系，学习时间和学习效果有些什么关系。如果我们把X变量和Y变量作图，X变量在x轴，Y变量在y轴。如果X变量增加，Y变量也增加，我们就可以得到一条直线，这就是线形回归，或是X对Y的简单回归。两个变量的回归并非那么绝对的，有程度之别，这就是它们之间的相关系数（correlation coefficient）。

11.5.1. 协方差

两个变量X和Y怎样一起变化，这就是协方差（covariance）的问题。协方差表示为COV（X，Y），其公式是：

$$COV(X, Y) = \frac{1}{n-1}\sum (X_i - \overline{X})(Y_i - \overline{Y}) \qquad (11.31)$$

假定我们有10个学生，分别参加英语科高考（X）和中学会考（Y），其成绩如下：

表11.24　　10个学生参加两个考试的成绩

学生	高考成绩（X）	会考成绩（Y）	$x-\overline{x}$	$y-\overline{y}$	
1	550	78	−21	−3.2	67.2
2	600	82	29	0.8	23.2
3	600	85	29	3.8	110.2
4	620	80	49	−1.2	−58.8
5	570	80	−1	−1.2	1.2
6	500	65	−71	−16.2	1150.2
7	500	75	−71	−6.2	440.2
8	630	90	59	8.8	519.2
9	610	89	39	7.8	304.2
10	530	88	−41	6.8	−278.8
平均值	571	81.2	$\sum=$	2278	
标准差	48.6369772	7.55424825		COV（X，Y）=	253.1111
				相关系数=	0.688896

两个变量的协方差是一个很有用的统计量，但是它作为一个表示线形关系的描写性变量却不大好理解，因为（1）它所使用的单位难以解释，因为所表示的其实是一个误差分数；（2）这个分数值是依赖于量表的，在表11.25里，不同的量表（如高考和会考）会产生不同的协方差。但是两个变量的关系并没有改变，因此我们通常使用不受量表影响的相关系数。

11.5.2. 积差相关系数

相关系数可以表示为：

$$r\ (X,\ Y)\ =\frac{COV\ (X,\ Y)}{S_X S_Y} \tag{11.32}$$

这个相关系数通常叫做 Pearson 积差相关系数。如果用原始分来计算，积差相关系数的公式是：

$$r=\frac{\sum XY-\ (\sum X)\ (\sum Y)\ /n}{\sqrt{\sum X^2-\ (\sum X)^2/n}\cdot\sqrt{\sum Y^2-\ (\sum Y)^2/n}} \tag{11.33}$$

一般的计算机编程都使用这个公式，因为它无需循环。还有一种简化的方法来计算协方差：

$$COV\ (X,\ Y)\ =\frac{\sum XY-n\ \overline{XY}}{n-1} \tag{11.34}$$

得出协方差后，再除以两个数组的标准差，也可求得相关系数。表 11.25 的两个数组的 SXY 为 465930，使用公式 11.34，465930－10 (571) (81.2) /10－1，也可得 253.1111。相关系数使用得非常广泛，一般的科学型计算器都有此计算功能。

表 11.25　　5%和 1%显著性水平的相关系数

自由度	5%	1%	自由度	5%	1%
1	0.997	1.000	24	0.388	0.496
2	0.950	0.990	25	0.381	0.487
3	0.878	0.959	26	0.374	0.478
4	0.811	0.917	27	0.367	0.470
5	0.754	0.874	28	0.361	0.463
6	0.707	0.834	29	0.355	0.456
7	0.666	0.798	30	0.349	0.449
8	0.632	0.765	35	0.325	0.418
9	0.602	0.735	40	0.304	0.393
10	0.576	0.708	45	0.288	0.372
11	0.553	0.684	50	0.273	0.354
12	0.532	0.661	60	0.250	0.325
13	0.514	0.641	70	0.232	0.302
14	0.497	0.623	80	0.217	0.283
15	0.482	0.606	90	0.205	0.267
16	0.468	0.590	100	0.195	0.254
17	0.456	0.575	125	0.174	0.228
18	0.444	0.561	150	0.159	0.208
19	0.433	0.549	200	0.138	0.181
20	0.423	0.537	300	0.113	0.148
21	0.413	0.526	400	0.098	0.128
22	0.404	0.515	500	0.088	0.115
23	0.396	0.505	1000	0.062	0.081

相关系数的计算并不难，但我们怎样凭 r 值来判断两个变量的相关性有多大呢?经常会人把相关系数的显著性意义和相关性的高低混为一谈。以表 11.25 的数组为例，自由度为 9，查表得悉 5%显著性水平的临界值为 0.602，而相关系数为 0.689，0.689>0.602，因此我们可以拒绝无差别假设，认为这个相关系数是有显著意义的。但是相关系数的显著意义和相关系数的意义是两回事。确定其显著意义仅是第一步，如果达不到显著性水平，我们就不能推翻无差别假设，再进一步推断。达到显著性水平后，我们才能说相关系数的高或低。由此可见，相关系数的显著性水平和样本大小有关，样本小，误差大，因此要求也高。但是样本数大不等于相关性就一定高，样本数大只能说提高我们评估相关性的把握程度。那么两个变量的相关程度又怎样看呢? Connolly & Sluckin (1957) 提出一个判断相关系数的程度的表（见表 11.26），可供参考。

表 12.26　　相关系数的相关程度

0.90—1.00	相关性很高；关系十分密切
0.70—0.90	相关性高；关系明显
0.40—0.70	相关性中等；有实质性关系
0.20—0.40	相关性低；有某种关系，但较小
0.20—低于 0.20	相关性甚微；关系甚小，可置之不顾

还有另外一种解释方法，就是用 r 的乘方来说明一个变量能够解释另一个变量差异的程度，例如 X 和 Y 的相关系数为 0.80，它的乘方是 0.64，我们可以说，“在 X 中所观察到差异可以说明 64%的 Y 的差异。”

11.5.3. 等级相关

表 11.27　　两个教师排列的相关

学生	教师 A 的排列	教师 B 的排列	$(R-S)^2$
A	1	1	0
B	2	3	1
C	4	9	25
D	3	2	1
E	5	7	4
F	6	4	4
G	8	8	0
H	6	5	1
I	9	6	9
J	10	10	0
			D=45
			rs=0.73

在上面讨论到的积差相关是以变量遵循正态分布为前提的，但是有些时候这个假定不一定能站得住，例如我们想了解学生两个考试成绩的相关，一个是 100 道选择题的考试，而另一个是分数只有 5 级的作文考试。后者就很难说是正态分布的。还有的时候，数据并非以分数的形式来表示，例如一个班的教师觉得他把学生按等级排列更能说明问题。这时我们需要计算等级相关（rank correlations），通常叫做 Spearman 等级相关系数。等级相关的公式是：

$$r_s=1-\frac{6D}{n(n^2-1)} \qquad (11.35)$$

$D=S(R_i-S_i)^2$，R_i 是 X 的等级，S_i 是

Y 的等级。例如表 11.28 是两个教师对 10 个学生成绩的排列，表面看来，等级相关的计算很简单，但是除非样本很少，否则计算等级也很花时间。而且有时我们会认为两个或更多的学生的等级差不多，实在分不出来，例如 A 是最好的，可是 B、C、D 都差不多，排不出来，我们就必须把他们所属的位置加起来，再除以 3，(2+3+4) /3=3，即 B、C、D 都是 3。

11.5.4. 线性回归

如果我们把表 11.25 的高考和会考成绩作图，图 11.19 就可见这两个数组存在一定的线性的关系，我们可从 X 来预测 Y，或 Y 来预测 X。这种关系表示为：Y=α+βX，α 是 Y 轴的截距 (intercept)，而 β 是这条线的斜率 (slope)，或称回归系数。只要我们求出 α 和 β，我们就可以确定这个回归方程式，并用它来进行预测。但是我们必须首先明确用什么来预测什么，这就牵涉到怎样确定两个变量的关系。这两个变量可分别称为自变量和依变量，在某种情况下，哪一个是自变量，哪一个是依变量，是带有任意性的，例如身高和体重的关系。我们可以用身高来预测体重，也可以反过来。在另外一些情况下，如年龄和外语学习的水平，则一般把年龄看成是自变量，而学习水平看成是依变量。还有一些情况则视实验或观察的目的而定，例如我们可以用会考来预测高考，也可以用高考来预测会考，但是比较多的情况是前者，因为会考在前，而且会考的重要性不如高考。不过这只是转换的公式不同，而 α 和 β 的计算公式都是一样的：

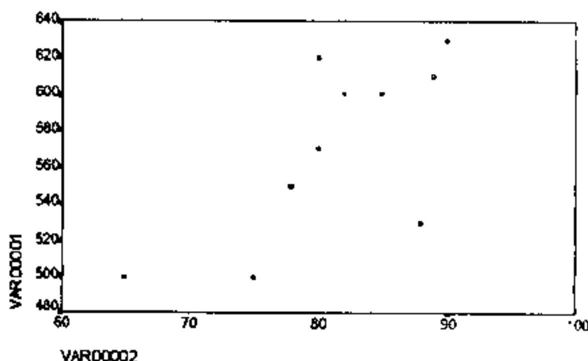


图 11.19

$$\beta=\frac{COV\ (X,\ Y)}{s_x^2} \qquad \alpha=\overline{Y}-\beta\overline{X} \qquad (11.36)$$

表 11.28　　根据会考成绩来预测高考成绩的结果

会考成绩	高考成绩	高考预测	残差	残差方
78	550	556.8069	−6.80685	46.33326
82	600	574.5483	25.45171	647.7897
85	600	587.8544	12.14564	147.5165
80	620	565.6776	54.32243	2950.926
80	570	565.6776	4.32243	18.6834
65	500	499.1472	0.852804	0.727274
75	500	543.5008	−43.5008	1892.318
90	630	610.0312	19.96885	398.7549
89	610	605.5958	4.404206	19.39703
88	530	601.1604	−71.1604	5063.808
截距=	210.8			11186.25
斜率=	4.44		s_r=	37.39361
			t=	2.68809

用表 11.25 的两组数据计算，β=0.10699，而 α=20.1046，回归方程式是：

$$\hat{Y}_i=20.1046+0.10699X_i$$

这是根据 X (高考) 来预测 Y (会考)；同样的，我们也可以用 Y 来预测 X：X= (Y−α) /β

$$\hat{X}_i=\frac{Y_i-20.1046}{0.10699}$$

所以我们可以根据一个学生的会考成绩来预测他的高考的成绩，例如一个会考拿到 80 分的学生

在高考里可能得到 80－20.1046/0.10699＝559.8。现把会考成绩做自变量，高考成绩作依变量，计算出其截距和斜率，并且作出高考的预测值，并计算出其残差。残差是观察值和预测值的差。

从理论上说，假定相关系数是 1，观察值和预测值应该是一样，其残差就是 0。换句话说，残差越大，相关性越低；残差越小，相关性就越高。残差的样本标准差（s_R）可以表示为：

$$s_r=\sqrt{\frac{\sum e_i^2}{n-2}} \tag{11.37}$$

在这里为 37.39。用这个标准差就可以计算出 t 值，其公式是：

$$t=\frac{\beta s_x\ \sqrt{n-1}}{s_r} \tag{11.38}$$

无差别假设为 b＝0，在 0.05％的显著性水平上自由度为 8 的临界值为 2.31，所得的 2.68<2.31，有显著性意义。这说明我们所拟合的线性回归模型是有价值的（不要忘记这两组数据的相关系数是 0.689，已经检验有显著性意义）。

接着下来要讨论的问题是预测值的置信区间的问题，既然两组数据的相关系数可高可低，因此所预测的值就会有不同的置信区间。假定我们根据会考 80 分预测高考的分数应为 566.7，这里可以有两种可能：一种可能是我们想了解 80 分的总体的平均分是否就是 566.7，其置信区间有多大，可使用下列公式：

$$\hat{Y}_X\pm\left\{C\times s_r\times\sqrt{\frac{1}{n}+\frac{(X-X)^2}{(n-1)\ s_X^2}}\right\} \tag{11.39}$$

C 是一个常数，为自由度（n－2）的 95％置信区间的临界值，查 t 概率分布表（附录 C）可知为 2.31。现预测值为 566.7，故 95％置信区间可计算为：

$$566.7\pm\left\{2.31\times37.39\times\sqrt{\frac{1}{10}+\frac{80-81.2)^2}{11\times7.55^2}}\right\}$$

或 566.7±27.6

因此我们可以估计会考的 80 分总体在 95％的情况下会在 539.1 到 594.3 之间。这种置信区间显示我们估算的精确程度。

也有另一情况，我们关心不是总体，而是具体的某一个拿了会考 80 分的学生的预测高考值 566.7 的置信区间。这时我们需要对公式（11.39）作修正如下：

$$\hat{Y}_X\pm\left\{C\times s_r\times\sqrt{1+\frac{1}{n}+\frac{(X-X)^2}{(n-1)\ s_X^2}}\right\} \tag{11.40}$$

结果是

$$566.7\pm\left\{2.31\times37.39\times\sqrt{1+\frac{1}{10}+\frac{(80-81.2)^2}{11\times7.55^2}}\right\}\text{或 }566.7\pm90.68$$

我们可以说，回归研究的结果表明，如果我们随机选一个会考 80 分的学生，他在 95％的情况下在高考中得分在 476 到 657 分之间。

11.5.5. 相关矩阵和部分相关

我们研究的往往不仅是一对变量的关系，而是若干个变量的关系。例如一个考试里有几道大题，这几个大题的关系我们都想观察，这就需要整理出一个相关矩阵：

表 11.29　　各大题的相关矩阵

	阅读	语法	综合填充	写作
阅读	1	0.8868	0.8537	0.8280
语法	0.8868	1	0.7159	0.8330
综合填充	0.8537	0.7159	1	0.6335
写作	0.8280	0.8330	0.6335	1

从矩阵来看各个大题之间都有较高的相关，阅读和语法的关系最为密切，而写作和综合填充则没有那么密切。假定我们想进一步了解语法和综合填充这两个变量和写作有些什么关系，那就需要计算部分相关。我们可以把写作定为 Y（依变量），而语法和综合填充分别定为 X_1 和 X_2（自变量）。X_1 和 X_2 的相关系数为 0.7159，说明它们之间关系密切，并非互相独立的。这两个变量一定包含了一些互相重叠的信息。现在我们想了解这两个变量和写作的关系，就必须在考虑其中一个变量（如综合填充）所提供的信息的前提下，分析另一个变量（如语法）和写作的关系，可表示为 $r(Y, X_1|X_2)$。标记的意思是：把 X_2 包含的信息考虑在内的 Y 和 X1 的相关。其计算的公式是（11.41）。

表 11.30　　一个考试的各大题分数

	阅读	语法	综合填充	写作
A	42.00	38.00	21.00	20.00
B	40.00	37.00	20.00	18.00
C	35.00	38.00	17.00	17.00
D	43.00	37.00	22.00	19.00
E	38.00	26.00	18.00	15.00
F	40.00	38.00	21.00	18.00
G	25.00	20.00	17.00	14.00
H	30.00	19.00	18.00	15.00
I	32.00	25.00	20.00	10.00
J	20.00	15.00	15.00	10.00

计算的结果为 0.704。对部分相关的解释和普通相关解释一样，$0.704^2=0.49=49\%$，这就是说，在考虑了综合填充以后，再把语法这个因素放到回归模型里可以解释所剩余的残差的 49%。

$$r(Y, X_1|X_2)=\frac{r(y, x_1)-r(y, x_2)\,r(X_1x_2)}{\sqrt{(1-r^2(y, x_2))(1-r^2(x_1x_2))}} \tag{11.41}$$

11.5.6. 相关系数的计算机运算

在一般的科学型的计算器里都有计算相关系数、截矩和斜率的功能，但在大型的统计软件包里，相关系数和回归的计算往往是分开的。例如在 SPSS 里，我们可以从主菜单到 Statistics/Correlate/Bivariate，去计算两个参数相关系数和相关距阵，到 Statistics/ Correlate/Partial 去计算部分相关。也可到 Statistics/Regression/linear 那里，既取得相关系数和，又取得回归方程式的有关数值。在 SPSS 里处理表 11.28 的结果显示如图 11.20：

```
* * * * * * * *    MULTIPLE    REGRESSION

End Block Number   1     All requested variables entered.
Equation Number   1     Dependent Variable..   VAR00002
Block Number   1.       Method: Enter          VAR00001
Variable (s) Entered on Step Number
   1..     VAR00001
Multiple R                .68890
R Square                  .47458
Adjusted R Square         .40890
Standard Error          37.39361

Analysis of Variance
                  DF     Sum of Squares     Mean Square
Regression         1     10103.74611        10103.74611
Residual           8     11186.25389        1398.28174

F=          7.22583          Signif F=0.0276
------------------ Variables in the Equation ------------------
Variable              B          SE B          Beta          T    Sig T
VAR00001        4.435358     1.650003      .688896        2.688  0.0276
(Constant)    210.848910   134.501086                     1.568  0.1556
End Block Number     1    All requested variables entered.
      * * * *   MULTIPLE    REGRESSION   * * * *
Equation Number 1     Dependent Variable..     VAR00002
Residuals Statistics:
Min        Max              Mean      Std Dev        N
*PRED         499.1472   610.0311   571.00   33.5058        10
*RESID        -71.1604       54.3224        0.0000     35.2550     10
*ZPRED      -2.1445          1.1649         0.0000      1.0000     10
*ZRESID       -1.9030    1.4527       0.0000        0.9428       10
Total Cases=        10
```

图 11.20

表 11.30 的数据可分别计算出相关矩阵和部分相关：

········· Correlation Coefficients ·········

	CLOZE	GRAMMAR	READING	WRITING
CLOZE	1.0000	.7159	.8537	.6335
	(10)	(10)	(10)	(10)
	P=.	P=.020	P=.002	P= .049
GRAMMAR	.7159	1.0000	.8868	.8330
	(10)	(10)	(10)	(10)
	P=.020	P=.	P=.001	P=.003
READING	.8537	.8868	1.0000	.8280
	(10)	(10)	(10)	(10)
	P=.002	P=.001	P=.	P=.003
WRITING	.6335	.8330	.8280	1.0000
	(10)	(10)	(10)	(10)
	P=.049	P=.003	P=.003	P=.

图 11.21

·········P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S·········

Controlling for.. CLOZE

	WRITING	GRAMMAR
WRITING	1.0000	.7024
	(0)	(7)
	P=.	P=.035
GRAMMAR	.7024	1.0000
	(7)	(0)
	P=.035	P=.

(Coefficient / (D.F.) / 2-tailed Significance)

图 11.22

11.6. 方差分析

在 11.4.3 里，我们谈到使用 t 检验的方法来观察两个样本的平均值的差异，还谈到两个以上的样本，虽然也可以用同样的检验方法，但会降低显著性水平。最好是使用方差分析（ANOVA）的方法。

方差分析是一种分析几个变量同时在起作用时的效应的统计程序。方差分析的目的是决定哪一个变量具有显著性效应，并且估算这些效应。它使用的方法是把观察中的总体的方差分为不同实验变量的方差和误差。这些观察通常是在实验环境下获得的，例如在 10.6.3 里，我们谈到了实验设计中的因子设计方法，这种方法在统计上就需要方差分析。农业中的作物试验田是最早使用方差分析方法的科学实验之一。实验设计需要规定观察的类型和数量，选择实验变量，考虑实验结构和使用随机化程序来避免结果的偏颇，其最终目的是以最小的花费获得最多的信息。如果我们在获得数据以前多花点功夫在设计实验上，就会取得更精确的结论。在我们设计好实验后，就可以提出描写实验的数学模型，根据数学模型进行合适的分析，即方差分析。

11.6.1. 单向方差分析

让我们先看一种最简单的方差分析——单向方差分析（One-way ANOVA），假定我们对三种不同的教学方法作教学实验，然后从参加三种方法的被试中随机抽 5 个人，取得他们的成绩的平均分来比较三种方法的差异。这样我们就可以分析两种不同的方差：一种是组间方差（between groups variance），即三种方法之间的差异。另一种方差是组内方差（within groups variance），组内方差又可理解为错误方差（error variance），因为这几个被试是随机抽样的，照理说不应有显著性差异，如果出现差异，那就是随机误差。

表 11.31　三种教学方法的单向方差分析

教学方法	I	II	III
	9	9	6
	12	10	8
	6	5	4
	8	5	5
15	11	7	
总计	50	40	30
平均值	10	8	6

首先我们要考虑的是自由度。自由度是 n－1，现有 15 个学生，自由度应为 14。但是我们有 3 个组，所以组间的自由度为 3－1＝2，而组内自由度为 12。

组间	2
组内	12
总计	14

然后计算下列各值：

（1）总计（GT，Grand Total.）＝120

（2）总平均（GM，Grand Mean.）＝8

（3）总计×总平均＝修正因素（c.f，correction factor）＝960

（4）总离差平方和（SS_{Total}，Total Sum of Squares.）$=9^2+12^2+\cdots+5^2+7^2-c.f.=132$

（5）组间离差平方和（SSBetween）$=50\times10+40\times8+30\times6-c.f.=40$

（6）组内平方和（SS_{Within}）＝总平方和—组间平方和＝132－40＝92

最后列表如下：

表 11.32　三种教学方法的反差分析表

方差源	离差平方和	自由度	均　方
组间	40	2	20
组内	92	12	7.67
总计	132	14	

（7）计算 F 值＝20/7.67＝2.73

（8）查附录 D 的 F 值分布表，0.05 显著水平的临界值为 3.88，0.01 显著水平的临界值为 6.93。所以

$$F=2.73<F_{(2,12,0.05)}=3.88$$

这三组的平均值没有显著性差异。查阅 F 值分布表需要按照组间和组内两个自由度来交叉找出临界值。

单向方差分析指的是按照一个范畴变量来比较几个组的平均值，在表 11.31 里，这个范畴变量是一种教学方法。在同一表里，我们也可以把范畴变量定为 5 个被试，比较他们的平均值 8，10，5，6，11 有没有显著性差异。计算方法是一样的，其结果是 3.61，但组间自由度为 4，而组内自由度为 10，查阅附录 D 的 F 值分布表得知 0.05 显著水平的临界值为 3.48，现 F＝3.61＞F（4，10，0.05）＝3.48。我们可以认为这几个被试的平均值有显著性差异。

单向方差分析的一般模型可表示为：

表 11.33　单向方差分析表中所用的公式

来源	自由度	SS	MSS	F-比例
组间	$m-1$	$SS_{Between}$	$s_b^2=SS_{Between}/m-1$	s_b^2/s_r^2
组内	$m(n-1)$	SS_{Within}	$s_r^2=SS_{Within}/m(n-1)$	
总计	$mn-1$	SS_{Total}		

在单向方差分析中，我们虽然通过 F 值来了解几个平均值在整体上有无差异，但是我们往往还想进一步了解那一对平均值有无差异，这时就需要作一些检验。检验的方法很多，结论都无大差异，在一些统计软件包里，我们可以自由选择。在这里我们只介绍其中一种，以见一斑。Tukey 检验法所设定的临界值为 $q_k\sqrt{s_r^2/n}$，在本例里，n＝5，而 q_k 的临界值是需要查表的，自由度为 12，n＝5 的 0.05 显著性水平的临界值为 4.51，因此 $4.51\sqrt{7.67/5}=5.59$。在这个基础上，我们再计算每组平均值之差，如 I 和 II 两种方法之差为 2，I 和 III 之差为 4，II 和 III 之差为 2，没有一组的平均值之差大于 5.59，因此它们之间都没有显著意义的差别。这种检验方法需要查表，而且又不是所有的统计学手册都有这种表，故我们现在一般都靠统计软件包来作。

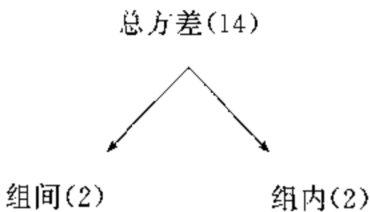


图 11.23

11.6.2. 重复测量设计

表 11.31 表示的是把三组被试随机地安排去参加三种不同教学方法的实验，然后比较平均值的差异。这是简单的单向方差分析。但也有另一种情况，即对一批被试作三次重复测量，然后比较其平均值有无差异。这就叫做重复测量设计（repeated measurement design）。简单的单向方差分析的方差可以表示为图 11.23。而重复测量设计则可表示为图 11.24。

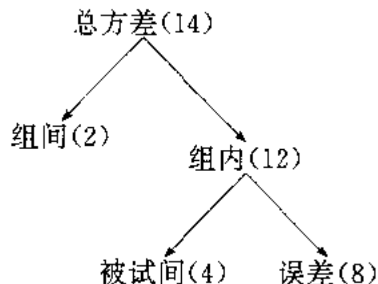


图 11.24

把图 11.31 和图 11.32 加以对比，我们可见重复测量设计多了一个层次，把组内方差再分为被试间和错误方差。重复测量设计的方差分析和简单的单向方差分析一样，需要多一个环节，计算被试间的 SS=24×8+30×10+…+33×11−c.f =78，在把组内 SS−78，便可得误差 SS。最后列表如表 11.35。从表中可见，组间 F 值 = 20/1.75 = 11.43，而被试间的 F 值 = 19.50/1.75 = 11.14。查 F 值分布表，F= 11.43 > $F_{(2,8,0.01)}$ = 8.65，我们可以说这三次重复测验的平均值在 0.01 显著性水平上是有差异的。F = 11.14 > $F_{(4,8,0.01)}$ = 7.01，我们也可以说，被试的平均值在 0.01 显著性水平上是有差异的。如果我们再把这个结果和简单的单向方差分析的结果来比较，就可以看出，原来是没有显著性差异的，而现在却有差异了。这显示了重复测验的一个特点，它可以使我们利用被试的系统性行为，使原来的误差减少，在两种计算中组间 MSS 都是 20，但简单的方差分析的误差（即组内方差）为 7.67，而重复测量的误差则只有 1.75，因此 F 值就从 2.73 跃为 11.43。

表 12.34　三次重复测量的设计

教学方法	I	II	III	总计	平均值
	9	9	6	24	8
	12	10	8	30	10
	6	5	4	15	5
	8	5	5	18	6
	15	11	7	33	11
总计	50	40	30		
平均值	10	8	6		

表 12.35　三次重复测量的方差分析结果

来源	自由度	SS	MSS
组间	2	40	20
组内	12	92	
被试间	4	78	19.50
误差	8	14	1.75
总计	14	132	

11.6.3. 双向方差分析

但是我们往往需要同时观察几个实验变量，例如社会语言学家想同时观察一个语言变量在语言环境和社会环境的使用情况，心理语言学家想观察不同类型的被试在不同语体中辨认词的反应时间，其目的都是想了解这些变量之间有些什么交互作用。在这种情况下，我们需要作因子实验设计，把不同的因子来进行分析，例如我们想比较不同地区、不同性别的学生在一个多项选择题的考试中的成绩（见 Woods，1986），可以有如下的设计：地区有四个水平（欧洲、南美、北非和东南亚），性别有两个水平（男性和女性）。

表 11.36　　40 个被试在一个多项选择题考试的成绩（按地区和性别分类）

	地		区		
性　别	欧洲（1）	南美（2）	北非（3）	东南亚（4）	总计
	10	33	26	26	
	19	21	25	21	
男性（1）	24	25	19	25	
	17	32	31	22	
	29	16	15	11	
小计	99	127	116	105	447
	37	16	25	35	
32	20	23	18		
女（2）	29	13	32	12	
	22	23	20	22	
	31	20	15	21	
小计	151	92	115	213	466
总计	250	219	231	231	913

我们可以用这一组数据来检验两个独立的无差别假设：在地区之间的平均值和在两种性别之间的平均值是否无差别。经过计算，我们得到下列结果（请参看下面 SPSS 的报告）：

表 11.37　　地区与性别的双向方差分析表

(a) 数据的主要效果的方差分析

来源	自由度	离差平方和	均方	F-比例
地区间	3	79.875	26.62	$F_{3,35}=0.54$
性别间	1	9.025	9.025	$F_{1,35}=0.184$
残差	35	1722.875	49.225	
总计	39	1811.775		

(b) 数据的主要效果和交互作用的方差分析

来源	自由度	离差平方和	均方	F-比例
地区间（L）	3	79.875	26.62	$F_{3,32}=0.68$
性别间（S）	1	9.025	9.025	$F_{1,32}=0.23$
交互作用（L×S）	3	384.875	128.29	$F_{3,32}=3.068$
残差	32	1338	41.813	
总计	39	1811.775		

把（a）和（b）来比较，我们可以看到（a）中的残差（1722.875）在（b）中分为两部分：交互作用（384.875）和残差（1338），然后再求出交互作用 F 的值（128.29/41.813）。

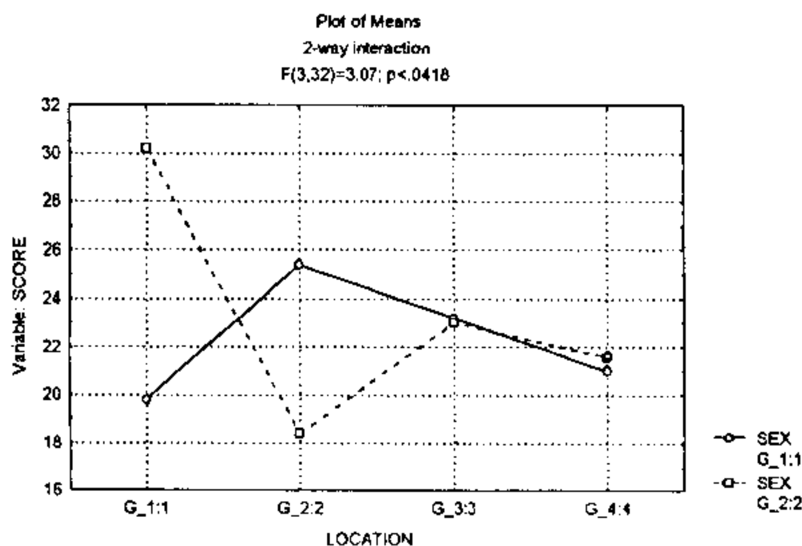


图 11.25 性别和地区的交互作用示意图

分析方差分析的结果可从两个角度来看：(1) 主要效果（Main effects），指一个独立于其他因素的因素（如地区、性别）内各个水平之间的差别，从（a）和（b）两个表看，F 值都比较小，可以说是没有显著性差异。(2) 交互作用（Interactions），指同时考察两个或两个以上因素时所出现的差别。现 L×S 的交互作用为 3.068，查 F 值分布表，自由度为 3 与 32，显著性水平为 0.05 时的临界值为 2.92，显著性水平为 0.01 时的临界值为 4.51，故我们可以说地区和性别的交互作用在 0.05 水平上有显著意义，在 0.01 水平上没有显著意义。交互作用可用图来表示：如果两个变量没有交互作用，则所作出的线基本上是平行的；如果两条线形成或最终形成一个交叉点，那就有交互作用。从图 11.23 可明显看出，在欧洲男学生的成绩不如女学生，但在南美则男学生倒过来比女学生的好。到了北非和东南亚，则男女生的成绩所差无几。

双向方差分析的一般模型可表示为：

表 11.38　　双向方差分析所用的公式

来源	自由度	离差平方和	均方	F-比例
地区间	L-1	SS_L	$MS_L=SS_L/(L-1)$	MS_L/MS_e
性别间	S-1	SS_S	$MS_S=SS_S/(S-1)$	MS_S/MS_e
交互作用	(L-1)(S-1)	SS_I	$MS_I=SS_I/(L-1)(S-1)$	MS_I/MS_e
误差（残差）	LS (n-1)	$SS_e=SS_T-SS_L-Ss_s-SS_I$	$MS_e=SS_e/(LS(n-1))$	
总计	N-1	SS_T		

11.6.4. 固定效果和随机效果模型

在上述例子里，学生是从四个不同的地区抽样而来的。统计结果告诉我们，地区并没有主要效果。如果我们暂时把交互作用放在一边，我们可以说，一个地区的学生的平均值和另一个地区的学生的平均值是很相似的。但是我们这个结论的应用范围有多宽？它是否只适用于这四个实际观察的地区？还是我们可以把结论延伸到别的地区？如果我们把结论只限于这四个地区，我们所作的分析是正确的。我们可以把这个模型称为固定效果模型（fixed effects models）。

如果我们把交互作用除外，这个模型可以表示为：

$$Y_{ijk}=\mu+L_i+S_j+e_{ijk}$$

Y_{ijk}表示在 i 个地区，属于 j 个性别的 k 个学生的分数，而 m 则表示总体的平均值。L_i 是 i 个地区的主要效果，而 Sj 则是 j 个性别的主要效果。e_{ijk}则是被试分数和总体平均值的随机误差。我们的无差别假设是：对四个地区而言，$H_0: L_i=0$；对两个性别而言，$H_0: S_j=0$。

如果我们想把这四个地区作为更多的地区的代表，把结论延伸得更加广泛，那就需要另外一种考虑和分析方法。我们必须建立一种机制把所观察到的地区的效果和别的不包括在实验里面的地区的效果联系起来。我们必须假定有许多地区，各自有自身的效果，L，对该地区的学生的平均值发生影响。而这些不同的 L 值是按正态分布的，其平均值为 0，而其标准差为 σ_L，那么我们关于地区的无差别假设是 $H_0: \sigma_L=0$。这样我们就可以假定，实验中的四个地区是许许多多地区中的随机的样本。这个模型可以称为随机效果模型（random effects models）。

还有一种混合模型（mixed models），即把有些变量看成是固定的（如性别），有些变量则是随机的（如地区）。固定效果模型和随机效果模型的 F 值计算视有无交互作用而略有不同。在教育研究里，经常使用的固定效果模型，随机效果和混合模型都是很罕见的，故怎样计算其 F 值，这里就略而不谈了。

11.6.5. 方差分析的计算机运算

在统计软件包里计算方差分析也很方便，但首先要注意数据的输入，以表 11.31 的三组数据为例，必须用下列方式输入。这是一个单向的方差分析，在 SPSS 里，应从主菜单到 Statistics /Compare Means/One-way ANOVA，然后把分数定为依变量，再选择一个因素来所方差分析，例如选“方法”，然后到 Post Hoc 那里选一种检验方法，来看这几种方法的平均值有无显著性差异。图 11.26 是选了 Tukey 检验法的结果。我们也可以选“被试”，得到的 F 值为 3.61，这里就不列出了。双向方差分析的数据输入方法也是一样的，但在 SPSS 里，就需要去 Statistics /ANOVA Models/Simple Factorial，然后还是一样指定依变量，再输入因素。和单向方差分析不同，这里可以决定多个因素，不但是双向，也可以是三向、四向……。在双向方差分析中，我们可以在子菜单的选择（Options）里决定是否需要把交互作用的方差计算出来。图 11.27 显示了表 11.36 在 SPSS 运算后的结果。细心的读者不难看出图 11.27 的结果和表 11.36 所列出的结果基本是相同的，但 SPSS 列出更多一个层次，它把地区、性别、交互作用的方差加起来，叫做能解释（explained）的方差，并计算其 F 值。

重复测量设计的输入数据办法不一样，因为重复测量的几个数组都是依变量，所有必须把它们并排在一起（如计算相关系数时那样），然后去 Statistics /ANOVA Models/Repeated Measures。进入子菜单后，必须先定义因素的水平（level），例如单向方差分析的例子重复测量 3 次，就必须定义为 3，然后再把分数定义为组内被试变量（within subjects variables），就可以得到我们在重复测量设计的例子里所得到的结果，限于篇幅，我们也不再把结果列出。

被试	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
方法	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
分数	9	12	6	8	15	9	10	5	5	11	6	8	4	5	7

```
---------- O N E W A Y ----------
      Variable SCORE
   By Variable METHOD
                              Analysis of Variance
                              Sum of      Mean        F          F
Source             D.F.       Squares     Squares   Ratio Prob.
Between Groups       2        40.0000     20.0000    2.6087     .1146
Within Groups       12        92.0000      7.6667
Total               14       132.0000
---------- O N E W A Y ----------
Variable SCORE
By Variable METHOD
Multiple Range Tests: Tukey—HSD test with significance level .050
      The difference between two means is significant if
      MEAN (J) —MEAN (I) >= 1.9579* RANGE* SQRT (1/N (I) + 1/N (J))
      with the following value (s) for RANGE: 3.77
      — No two groups are significantly different at the .050 level
      Homogeneous Subsets (highest and lowest means are not significantly different)
Subset 1
Group     Grp 3        Grp 2        Grp 1
Mean      6.0000       8.0000       10.0000
```

图 11.26

* * * A N A L Y S I S O F V A R I A N C E * * *

SCORE
by LOCATION
SEX
UNIQUE sums of squares
All effects entered simultaneously

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	88.900	4	22.225	.532	.713
LOCATION	79.875	3	26.625	.637	.597
SEX	9.025	1	9.025	.216	.645
2-Way Interactions	384.875	3	128.292	3.068	.042
LOCATION SEX	384.875	3	128.292	3.068	.042
Explained	473.775	7	67.682	1.619	.166
Residual	1338.000	32	41.813		
Total	1811.775	39	46.456		

40 cases were processed.
0 cases (.0 pct) were missing.

图 11.27

11.7. 多元分析方法

在上述统计分析中，我们讨论的都是一些最基本的方法，主要是处理的几个变量之间的关系，有的是差异的关系（如 t 检验和方差分析），有的是联系的关系（如相关和回归）。这些统计方法可以称为单维分析（Univariate Analysis）。但在现实生活里，我们面对的是更为复杂的关系，如一群被试中的每个人都有几个变量，而一个变量的值和另一个变量的值有时很接近，有时则无大关系。多元分析（Multivariate Analysis）方法研究的是同一个人或物的几个方面（维度）的关系，例如一件商品有几个不同价格，一个被试对几种显示有不同反应时，一个机体有几个生物指标，一个国家有几个社会和经济指标等等。和单维分析一样，我们必须假定多维观察的也是一个总体的随机样本，但是这些样本的几个因素之间具有某些依存或相关的关系，多元分析的任务就是揭示它们的关系，使我们更深刻地理解这种联合变异（joint variation）的复杂性。多元分析也有人称为多维测量（Multidimensional Measurement）。

多元分析既可以用来研究变量之间的关系，例如主要成分分析法观察的是变量之间的相关程度；也可以用来研究个体（或群体）之间的关系，例如判别分析观察的是一个数据矩阵

里的个体按照其变量值应该属于这个组别还是另一个组别。多元分析的数据集合都比较大，例如 50 个个体的 10 个变量就会有 500 个数值。这就需要把数据进行提炼，一种方法是使用归纳法，编制汇总表；另一种办法是减少那些关系不大或多余的变量（如多元回归分析）；还有一种办法是使用一些数量不多的“推导变量”，以归纳出数据集合的主要特征。所以在某个意义上说，多元分析的目标是在复杂的多元世界里归结为更简约的型式，去芜存精，发现内部规律。

多元分析牵涉到多因素之间的互相关系，它的数学基础是线性代数，特别是矩阵代数，和由此生成的相关矩阵。在下面的讨论中，我们有意避开这些专业性知识，而且因为数据众多，运算繁复，通常都要用计算机来完成，所以不作具体的运算示范。我们只集中介绍基本概念和使用范围，其结果则用统计软件包来计算，并不另辟专门章节。

11.7.1. 多元回归分析

多元回归分析 (Multiple Regression Analysis) 虽然是一种多元分析，但它往往被放在单元回归分析中一起讨论，因为（1）它的回归模型建筑在单元概率分布模型，模型的估计往往需要用到一些单元的理论分布模型，如正态和 t 分布。而其他的多元分析往往需要多元正态分布模型。（2）在多元回归分析里，数据矩阵中的一个变量被指定为依变量，它和其他的变量都是对称的；在别的多元分析里，所有的变量都是互相对称的。

我们之所以到现在才介绍多元回归分析，主要是为了叙述的方便，因为多元回归分析往往要先做方差分析，需要掌握它的基本概念，所以最好在分别介绍了单元回归分析和方差分析以后，再来谈多元回归分析，可以事半功倍。而且在进入其他多元分析前，谈多元回归分析是一个很好的过渡。但是读者需要了解它和其他多元分析的异同。

在 11.5.4 里，我们谈到线性回归的方程式 Y＝a＋bX ，这个方程式可以扩展为：

$$Y_i = \alpha + \beta X_{2i} + e_i$$

e_i 是一个被试的真正分数和他的预测的分数之间的残差。这个公式表示的是一个自变量和一个依变量之间的关系。但是在现实生活里，我们碰到的往往是几个自变量和一个依变量的关系，例如阅读理解能力是我们的依变量，而影响这个依变量的是许多因素，如单词的常用度、单词的长度、句子的复杂度、句子的长度、文章的语体、文章的熟悉度等等。这些因素各自对阅读理解能力都有所影响，但把它们联合起来又是怎样影响的，我们就必须把这些因素组合到多元回归的方程式里：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i,\ i = 1,\ 2,\ \cdots,\ n \qquad (11.42)$$

让我们看一个简单的例子 (Woods，1968)，依变量 Y 是水平考试，我们想了解综合填充 (X_1) 和词汇 (X_2) 两个因素组合在一起和水平考试的关系，可以先计算一个相关矩阵。从矩阵的结果看来水平考试和综合填充、词汇都有较高的相关。如果单拿综合填充来计算线形回归，可有下列结果：

表 11.39 综合填充、词汇和水平考试的相关距阵

	水平考试	综合填充	词汇
水平考试	1	0.8963	0.8378
综合填充	0.8963	1	0.7082
词汇	0.8378	0.7082	1

多元相关=0.896，标准误为 7.372，回归公式为：

$$Y=24.19+3.433X$$

综合填充的平均值为 10.87，标准误为 4.26。根据（11.40）公式计算综合填充为 8 分的预测 Y 值的置信区间为

$$\hat{Y}\ (8)\ \pm\left\{2.04\times7.372\sqrt{1+\frac{1}{30}+\frac{(8-10.87)^2}{29\times2.26^2}}\right\}=51.65\pm15.40$$

表 11.40 综合填充和词汇对水平考试的多元回归

学生	水平考试 (Y)	综合填充 (X_1)	词汇 (X_2)	学生	水平考试 (Y)	综合填充 (X_1)	词汇 (X_2)
1	93.00	19.00	29.00	16	78.00	13.00	28.00
2	86.00	16.00	26.00	17	75.00	15.00	24.00
3	69.00	14.00	25.00	18	70.00	16.00	30.00
4	80.00	11.00	25.00	19	52.00	11.00	16.00
5	92.00	19.00	31.00	20	67.00	14.00	19.00
6	53.00	7.00	20.00	21	45.00	5.00	17.00
7	55.00	6.00	19.00	22	40.00	5.00	16.00
8	72.00	13.00	23.00	23	36.00	4.00	13.00
9	79.00	16.00	32.00	24	49.00	9.00	22.00
10	45.00	4.00	15.00	25	67.00	13.00	21.00
11	41.00	7.00	15.00	26	55.00	8.00	28.00
12	51.00	9.00	19.00	27	36.00	7.00	9.00
13	60.00	10.00	16.00	28	58.00	9.00	20.00
14	72.00	11.00	31.00	29	69.00	14.00	21.00
15	42.00	9.00	10.00	30	58.00	12.00	18.00

这就是说在 36 到 67 分之间。这个区间太广，用处不大，这主要是因为标准误太大，必须拟合另一个更好的预测模型。如果我们把词汇的分数也考虑在回归模型里，会不会有所改善呢？这个新的模型可以写为：

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+e_i$$

于是我们需要估计三个参数 β_0（即 α，亦称常数），β_1 和 β_2，使用 SPSS 所得到的结果表示为图 11.28。从图中可见，除了回归分析外，统计程序首先还做一次方差分析，这是因为线性回归分析的目的是估计两个变量之间的线性关系的参数，利用这种线性关系我们就可以进行预测。如果两个变量是互相独立的，要估算它们之间的线性关系就不合适了。方差分析是线性回归

的显著性检验，它所得的F值可以用来检验应否拒绝无差别假设。现F值为104.96937，有非常显著的意义，说明两个变量并非相互独立的。这样我们才能进一步使用回归公式的参数。我们可见多元相关从0.896提高到0.941，而标准误则从7.372降低到5.72。回归方程式就成

```
* * * * M U L T I P L E   R E G R E S S I O N * * * *
Listwise Deletion of Missing Data
Equation Number 1      Dependent Variable..    PROF
Block Number 1.     Method: Enter       CLOZE       VOC

Variable (s) Entered on Step Number
  1..      VOC
  2..      CLOZE

Multiple R              .94130
R Square                .88605
Adjusted R Square       .87761
Standard Error         5.71529

Analysis of Variance
               DF      Sum of Squares      Mean Square
Regression      2        6857.55684        3428.77842
Residual       27         881.94316          32.66456

F=        104.96937       Signif F =   .0000

----------------- Variables in the Equation -----------------

Variable          B           SE B      95% Confdnce Intrvl B        Beta

CLOZE        2.328706       .352480    1.605478    3.051934     .607933
VOC          1.057377       .238938     .567116    1.547638     .407211
(Constant)  13.707845      3.743343    6.02714021.388550
------------ in ------------
Variable          T Sig T
CLOZE           6.607  .0000
VOC             4.425  .0001
(Constant)      3.662  .0011

End Block Number 1 All requested variables entered.
```

图 11.28

为

$$\hat{Y}=13.708+2.329X_1+1.057X_2$$

增加了词汇这个项目以后，回归模型可以解释水平考试分数中更多的变异，减少了残差。表11.41进一步说明在95%置信度下对水平考试的预测区间，单独使用 X_1 或 X_2 的置信区间都比较宽，而共用两个测试后的置信区间都要窄一些，即准确程度要提高一些。

表 11.41　　预测学生的水平考试的分数的95%置信区间

综合填充 (X_1) 分数	词汇 (X_2) 分数	自变量		
		只用 X_1	只用 X_2	X_1 和 X_2 共用
5	15	26—57	29—67	29—54
5	20	26—57	40—78	34—59
5	25	26—57	51—89	39—65
10	15	44—74	29—67	41—65
10	20	44—74	40—78	46—70
10	25	44—74	51—89	51—75
15	15	61—91	29—67	51—77
15	20	61—91	40—78	57—82
15	25	61—91	51—89	63—87

多元回归分析牵涉到几个因素对依变量的联合作用，但是这几个因素的作用不可能是完全一样的，而且因素之间不可能是毫无关系的。这就出现所谓多元共线 (multicollinearity) 的问题。让我们先看一组数值（表11.42），它们的相关矩阵表示如表11.43：

表 11.42　　三个大题和水平考试的分数

学生	水平考试 (Y)	阅读 (X_1)	语法 (X_2)	综合填充 (X_3)
1	90.00	19.00	19.00	17.00
2	85.00	16.00	18.00	18.00
3	80.00	13.00	17.00	15.00
4	86.00	14.00	16.00	13.00
5	70.00	12.00	15.00	12.00
6	77.00	10.00	10.00	13.00
7	78.00	12.00	15.00	14.00
8	60.00	11.00	11.00	11.00
9	65.00	13.00	12.00	12.00
10	50.00	10.00	10.00	10.00

表 11.43　三个大题和水平考试的相关距阵

	水平考试	阅读	语法	综合填充
水平考试	1.0000	0.7280	0.8127	0.8495
阅读	0.7280	1.0000	0.8601	0.8126
语法	0.8127	0.8601	1.0000	0.8429
综合填充	0.8495	0.8126	0.8429	1.0000

从相关矩阵看来，几个测试和水平考试的相关性都较高，较高的是综合填充（0.8495），而且它和其他几项相关也高。阅读和语法的相关为 0.8601，是所有相关系数中最高的，但我们现在要预测的是水平考试，这就使我们怀疑多元共线的存在。使用 SPSS 来计算把全部自变量考虑在内的线性回归，可用 ENTER 的选项，得到的结果是：

Multiple R　0.86936　Standard Error　7.68479
R Square　0.75578　F=6.18940　Signif F=0.0288
Adjusted R Square　0.63367
回归方程式为：$\hat{Y}=18.25+(-0.41)X_1+1.475X_2+2.96X_3$

如果我们想检验是否有多元共线的问题，可去 STEPWISE 的选项，运算的结果告诉我们：

Multiple R	0.84953	Standard Error		7.10444	
R Square	0.72170	F=20.74602		Signif F=0.0019	
Adjusted R Square	0.68691				
··········· Variables in the Equation ···········					
Variable	B	SE B	Beta	T	Sig T
Cloze	4.230769	0.928864	0.849530	4.555	0.0019
(Constant)	16.984615	12.739323		1.333	0.2192
··········· Variables not in the Equation ···········					
Variable	Beta In	Partial	Min Toler	T	Sig T
Reading	0.110851	0.122467	0.339683	0.326	0.7536
Grammar	0.333626	0.340307	0.289557	0.958	0.3702

图 11.29

这就是说，把阅读和语法这两个因素排除在外，多元相关并没有多大降低，而 F 值却提高了，实际上用一个因素（综合填充）也能取得满意的结果，而且可以减少运算程序。线性回归方程式为：

$\hat{Y}-19.98+4.23X_3$

11.7.2. 多元方差分析

多元方差分析 (Multivariate Analysis of Variance，MANOVA) 可以有两个或多个依变量，在这个意义上我们可以说单元方差分析 (Univariate ANOVA) 是一种特殊的多元分析，因为它只有一个依变量。单元方差分析可以是单向的，也可以是双向的或多向的。以表 11.44 为例，如果我们的实验设计是在两种不同类型的学校里实验三种方法，依变量是考试成绩，这是 2×3 的双向因子设计。但如果我们在设计里增加一个依变量，每一种方法的实验前后都组织一次考试，这成了多元方差分析。

表 11.44　　两类学校三种方法在实验前后的分数

	方法 1		方法 2		方法 3	
	前	后	前	后	前	后
非重点学校	9	20	9	10	6	6
	12	24	10	22	8	8
	6	12	5	12	4	4
	8	12	5	10	5	5
	15	25	11	20	7	7
平均值	10	18.6	8	14.8	6	6
重点学校	10	21	11	11	7	10
	10	22	12	13	8	10
	12	24	9	10	9	10
	14	25	9	10	10	11
	15	25	14	14	11	12
平均值	12.2	23.4	11	11.6	9	10.6

如果我们使用双向方差分析，对方法实验前的考试成绩而言，三种方法和两种学校的 F 值分别为 5.545 和 9.551，均有非常显著性的差异。方法和学校之间没有交互作用，F 值只有 0.091。但对方法实验后的考试成绩而言，三种方法的 F 值为 29.691，有非常显著性的差异；两种学校的 F 值却只有 2.318，没有显著性差异。方法和学校之间有交互作用，F 值为 3.766。

使用多元方差分析实际上是把这些结果综合起来考察，图 11.30 是 SPSS 计算出来的结果。在进行方差分析前，必须进行多元显著性意义检验分析，SPSS 使用了好几种统计量，其结果大致一样，这些检验必须有显著性意义，如图 11.30 的 Pillais F 值为 4.15，显著性水平为 0.006，Hotellings F 值为 5.74，显著性水平为 0.001，Wilks F 值为 4.96，显著性水平为 0.002。检验表明有显著性意义，就可继续观察下面的单元 F 检验，方法试验前的考试成绩 F 的值为 0.091，无显著性差异，但实验后的考试成绩的 F 值为 3.765，显著性水平为 0.038，有交互作用。在 SPSS 里进行多元方差分析，应从主页到 Statistics/Anova Models/Multivariate，到 Multivariate 后，必须先指定依变量，然后指定因素。

```
* * * * * * A n a l y s i s o f V a r i a n c e — design 1 * * * * * *
EFFECT .. METHOD BY SCHOOL
Multivariate Tests of Significance (S=2, M=-1/2, N=10 1/2)
Test Name        Value    Approx. F Hypoth. DF    Error DF   Sig. of F
Pillais        0.51437    4.15481      4.00       48.00     0.006
Hotellings     1.04389    5.74137      4.00       44.00     0.001
Wilks          0.48807    4.96104      4.00       46.00     0.002
Roys           0.50958
Note.. F statistic for WILKS' Lambda is exact.
------------------------------------------------------------------
EFFECT .. METHOD BY SCHOOL (Cont.)
Univariate F-tests with (2, 24) D. F.
Variable   Hypoth. SS    Error SS Hypoth. MS    Error MS       F      Sig. of F
BEFORE       1.06667    140.80000    0.53333     5.86667    0.09091    0.913
AFTER      104.06667    331.60000   52.03333    13.81667    3.76598    0.038
------------------------------------------------------------------
```

图 11.30

一般来说，如有交互作用，最好有附图。我们用 Statistica 对两组分数的交互作用作图，图 11.31 表示不同类型学校的三种方法在试验前的考试分数，没有交互作用，故两条线（代表学校）的发展基本上是平行的，不会相交。图 11.32 则不同，表示实验后的考试分数在两类学校三种方法是不一致，两条线有交互作用。

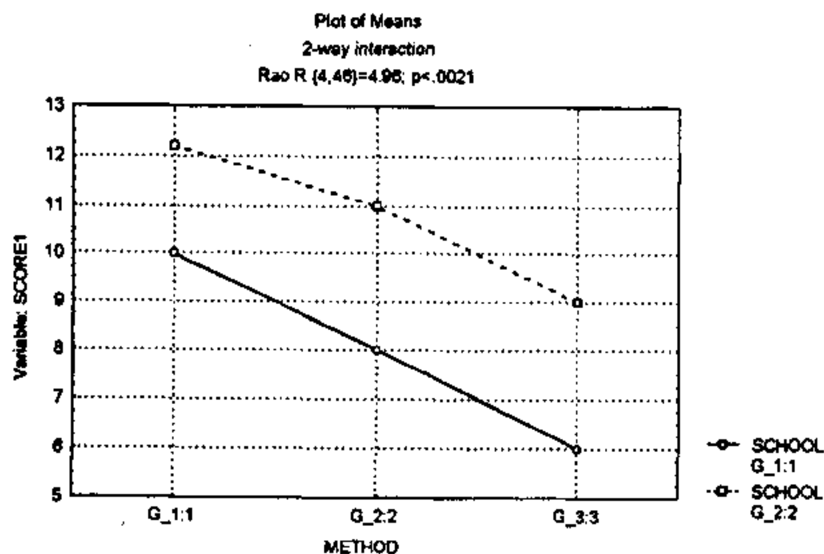


图 11.31　两类学校三种方法实验前的分数：无交互作用

11.7.3. 主要成分分析和因子分析

在多元回归分析中，我们谈到多元共线的问题，而解决之道是在回归公式中减少变量数目。主要成分分析(Principal Components Analysis）和因子分析(Factor Analysis）的出发点也是在分析中减少因素，或是找寻一些支配几个变量的共同因素。当几个变量有很多共同的特征时，我们就可以说存在一个支配这些变量的共同因素。例如我们测量一群人的手臂、腿和身体的长度，发现它们有明显的相关，一个个子高的人会有长手臂和长腿。这些相互关系就构成了一个共同因子：线性体积。如果我们把眼睛颜色和使用左手的情况和三种长度一起来计算相关，就会发现它们和长度没有什么关系，因此不属于同一个因子。因子分析方法就是利用相关矩阵所提供的数据，把相互联系的事物的关系加以系统化、有序化，找出一些支配变量的共同因子。因此因子分析方法有助于我们描写一个群体：我们可以对群体进行各种测量，了解其相关，而最后概括出足以描写这个群体的几个因子或维度。

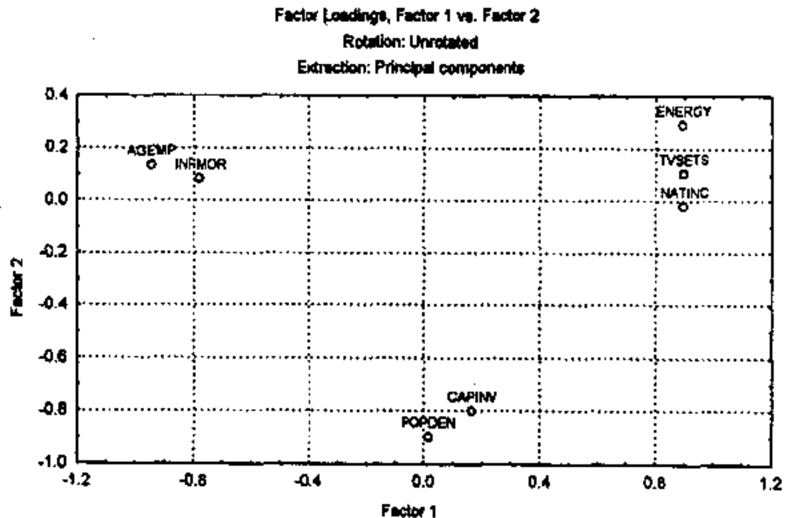


图 11.32 两类学校三种方法实验后的分数：有交互作用

11.7.3.1. 相关系数的几何法

相关系数可以用几何法来表示，两个变量的相关性就是两条直线之间的夹角。这些直线称为矢量(vectors)：

$$\cos\theta=\frac{AC}{AB}$$

$$AB=1$$

$$AC=correlation$$

$$\cos 30°=0.866$$

$$\cos 45°=0.707$$

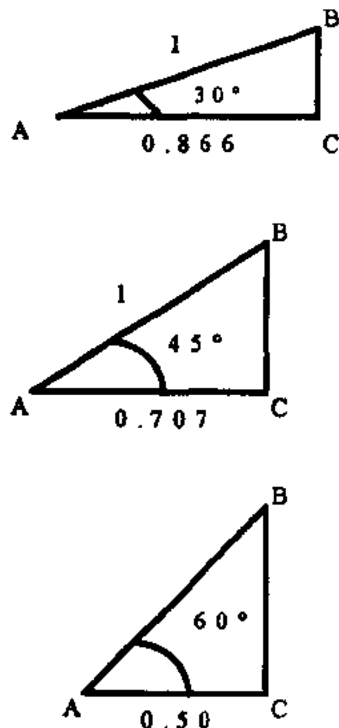


图 11.33 相关系数的几何表示法

cos60°＝0.50

cos90°＝0

cos0°＝1

从图 11.33 可见，相关系数可用 cosθ 来表示：如果 AC 与 AB 完全重叠，cosθ 便是 1，即绝对的相关；如果 AC 和 AB 成直角，cos90°0，即毫无相关。

假定我们有三个变量，A，B，C，这三个变量的相互关系可表示为一个相关矩阵，右上角为相关系数，而左下脚为 θ：

表 11.45　三个变量的相关距阵和 θ

	变量 A	变量 B	变量 C
变量 A	1.000	0.5000	0.3420
变量 B	50°	1.000	0.7660
变量 C	70°	40°	1.000

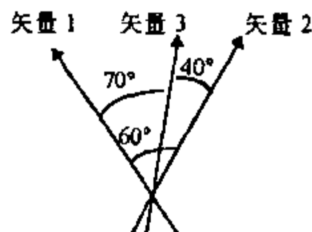


图 11.34

我们用矢量 1，2，3 来代表这三个变量，其相互关系可以表示为图 11.34。

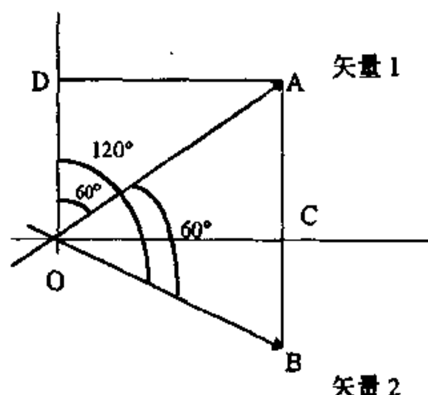


图 11.35

从图 11.34 可见，这三个变量是互相联系的，如果我们增加一个新的矢量，作为这三个矢量的参照线，找出这三个矢量和参照线夹角的 cos，我们就可找出一个足以说明这三者关系的一个共同因子。这好比一把半开的伞，向四周散开的伞架就是各个变量的矢量，而伞柄就是居中参照矢量，和各个矢量形成一定关系。这个参照矢量就是因子分析中的因子。假定我们有两个矢量 OA 和 OB，它们之间的夹角是 60°，我们可以增加一对矢量，如图 11.35：

cosAOB＝cos60°＝0.5

cosAOC＝cosBOC＝cos30°＝0.866

cosDOA＝cos60°＝0.5

cosDOB＝cos120°＝－0.5

11.7.3.2. 形心法

OD 和 OC 是我们增加的一对直角的参照矢量，根据 OA 和 OB 和参照矢量的关系，我们可分离出两个因子，OA 与 OD 的夹角为 60°，故 cos60°的因子负荷为 0.5；OB 与 OD 的夹角为 120°，故 cos120°＝－cos60°＝－0.5；OA 与 OC 的夹角为 30°，OB 与 OC 的夹角也是 30°，因此∠AOC 与∠BOC 的因子负荷均为 cos30°＝0.866。这样我们就可以整理出表 11.46。

每个变量的因子负荷的平方和都是 $0.866^2+0.5^2=1$；每个因子的负荷也可计算其平方和，这就是它的特征值，所以因子 1 的特征值为 $0.866^2+0.866^2=1.5$，因子 2 的特征值为 $0.5^2+0.5^2=0.5$。最后还可算出每个因子负荷所占的比例，因子 1 为 0.75，因子 2 为 0.25。

这个意思是：我们分离出两个因子，第一个因子可解释 75%的方差，第二个可解释 25%的方差。第一个因子所能解释的方差是第二个因子的三倍，故第一个因子是主要的。这就是主要成分分析法的理论模型。当然在实际计算中，变量的平方和加起来不会是 1，而因子的百分比加起来也不会是 100，因为任何的试验或测试都不能解释全部的方差。用上述分离因子和它们所占的百分比的方法叫做形心法（The Centroid Method）。

表 11.46　两个变量的因子负荷和特征值

	因子 1（OC）	因子 2（OD）	每个变量负荷的平方和
变量 1	30°（0.866）	60°（0.5）	1
变量 2	30°（0.866）	120°（−0.5）	1
	0.866^2	0.5^2	
	0.866^2	-0.5^2	
eigenvalue（特征值）	1.5	0.5	
百分比（eigenvalue/N×100）	75	25	

11.7.3.3. 主要成分分析

让我们先看一个具体的资料，OECD（经济合作与发展组织）于 1981 年根据人口密度（POPDEN）、农业雇员比例（AGEMP）、国民收入（NATINC）、机器投入（CAPINV）、婴儿死亡率（INFMOR）、能源（ENERGY）、每百人的电视机数（TVSETS）等项目公布了它的 14 个成员国的资料，如下表：

表 11.47　OEDC 成员国经济和社会特征

	POPDEN	AGEMP	NATINC	CAPINV	INFMOR	ENERGY	TVSETS
澳洲	2.00	6.00	8.40	10.10	12.00	5.20	36.00
法国	97.00	9.00	10.70	9.20	10.00	3.70	28.00
西德	247.00	6.00	12.40	9.10	15.00	4.60	33.00
希腊	72.00	31.00	4.10	8.10	19.00	1.70	12.00
冰岛	2.00	13.00	11.00	6.60	11.00	5.80	25.00
意大利	189.00	15.00	5.70	7.90	15.00	2.50	22.00
日本	311.00	11.00	8.70	10.90	8.00	3.30	24.00
新西兰	12.00	10.00	6.80	8.00	14.00	3.40	26.00
葡萄牙	107.00	31.00	2.10	5.50	39.00	1.10	9.00
西班牙	74.00	19.00	5.30	6.90	15.00	2.00	21.00
瑞典	18.00	6.00	12.80	7.20	7.00	6.30	37.00
土耳其	56.00	61.00	1.60	8.80	153.00	0.70	5.00
英国	229.00	3.00	7.20	9.30	13.00	3.90	39.00
美国	24.00	4.00	10.60	7.30	13.00	8.70	62.00
平均值	103	16	7.7	8.2	25	3.8	27
标准差	101	16	3.6	1.5	38	2.2	14

根据这个表，我们可以计算出其相关矩阵：

表 11.48 **OEDC 成员国经济和社会特征的相关距阵**

	POPDEN X_1	AGEMP X_2	NATINC X_3	CAPINV X_4	INFMOR X_5	ENERGY X_6	TVSETS X_7
X_1	1.00	−0.15	0.02	0.49	−0.13	−0.25	−0.07
X_2	−0.15	1.00	−0.79	−0.18	0.89	−0.72	−0.78
X_3	0.02	−0.79	1.00	0.19	−0.60	0.83	0.72
X_4	0.49	−0.18	0.19	1.00	0.00	0.01	0.13
X_5	−0.13	0.89	−0.60	0.93	1.00	−0.49	−0.52
X_6	−0.25	−0.72	0.83	0.01	−0.49	1.00	0.92
X_7	−0.07	−0.78	0.72	0.13	−0.52	0.92	1.00

从矩阵中可见，有些变量是有密切关系的，如能源消耗与电视机数的相关为 0.92，与国民收入的相关为 0.83。农业雇员与婴儿死亡率的相关为 0.89，与电视机数的相关为－0.78，与能源消耗为－0.72，等等。但是有些变量的关系又不明显。主要成分分析和因子分析企图把这些复杂的关系简约化。

```
-------------- F A C T O R  A N A L Y S I S --------------
PC     extracted     2 factors.
Factor Matrix:
                     Factor 1        Factor 2
AGEMP                -.94430         -.13543
CAPINV                .16475          .79883
ENERGY                .89317         -.29305
INFMOR               -.78141         -.08354
NATINC                .89678          .01816
POPDEN                .01805          .89606
TVSETS                .89717         -.10645
Final Statistics:
Variable          Communality  *  Factor   Eigenvalue   Pct of Var   Cum Pct
AGEMP                 .91005   *     1       3.93668       56.2        56.2
CAPINV                .66527   *     2       1.56392       22.3        78.6
ENERGY                .88363   *
INFMOR                .61759   *
NATINC                .80455   *
POPDEN                .80325   *
TVSETS                .81624   *
```

图 11.35

在 SPSS 里，我们从主菜单到 Statistics/Data Reduction/Factor，就可以进行主要成分分

析和因子分析。在 Factor Analysis 的菜单里，我们可以到 Extraction 选择 Method，如果不选择，SPSS 就自动进入主要成分分析。一般来说，程序还能以特征值为 1 去决定能分离出多少个因子。结果告诉我们，尽管有 7 个变量，但我们可以归结为两个因素，这两个因素的特征值 3.94 和 1.56，前者的百分比为 56.2%，后者为 22.3%，积累百分比为 78.2%。这就是这些变量的相互关系所能够说明的方差。第一个因子是主要的，它说明了 56.2%。第一个因子可以说是经济发展，它包括了国民收入、能源消耗、电视机数、农业雇员和婴儿死亡率（暂以 0.70 为标准①），而前三者的系数是正值，说明工业发达国家在这三方面都有较高的数值，后两者为负值，表明相反方向。第二个因子包括人口密度和机器投入，为什么它们归为一个因子，较难解释，无非是数学运算的结果。主要成分法把不算进主要因子里的其他变量都归入另一个因子。这两个因子里各个变量之间的关系可以作图，更为清楚，图 11.36 是使用 STATISTICA 作出的一个两维图。在图中第一个因子的 5 个变量都在一个方向，但按正负值的不同又分到两个角落。

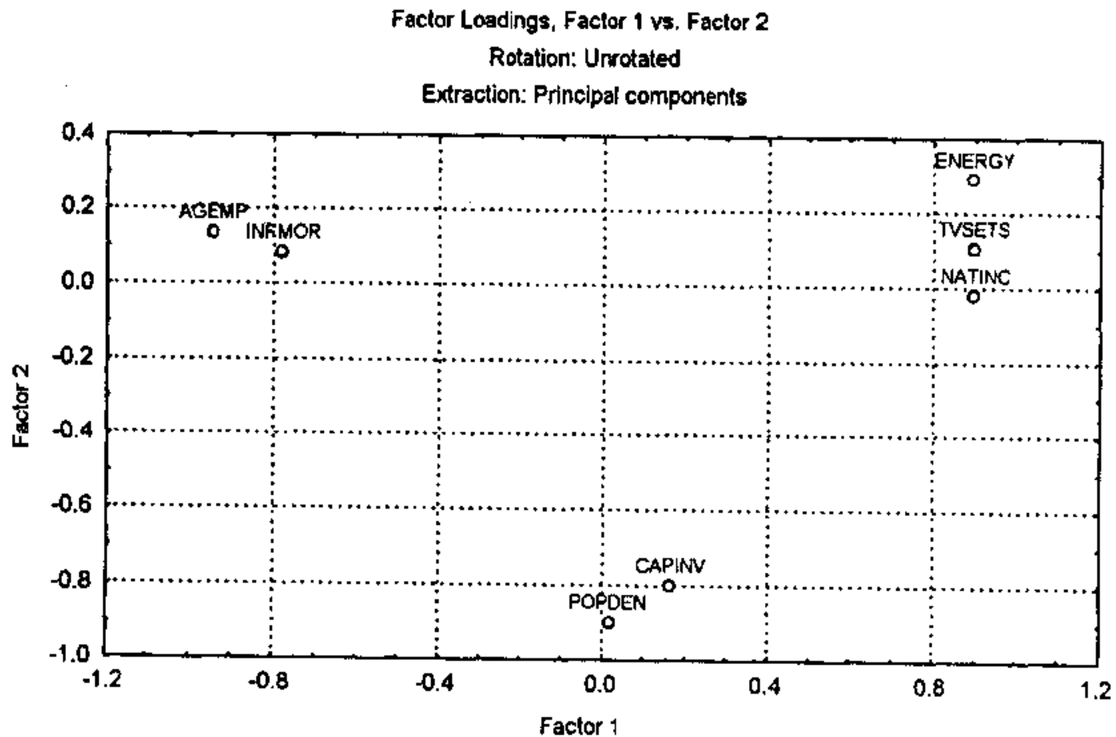


图 11.36 OEDC 经济和社会特征的因子分析图

我们还可以在软件里规定按国家列出其因子数值（factor scores），根据各个国家在这两个因子中的数值，就可以作出一个表示它们关系的图（见图 11.37）。

从这个例子看来，两个成分就能解释很大一部分的方差，这样一来就可以大大减少维度。所以这是一种很有用的探索性工具，帮助我们去了解一些处于多维度状态下的结构。例如我

① 选多大的因子负荷作为决定一个变量是否属于一个因子有很大的伸缩性。选 0.70 作为标准，意味着一个变量在一个因子中能解释 $0.70^2 \times 100 = 49\%$ 以上的方差。Child（1970）认为，如果样本数大于 50，能解释 10% 方差的就可以作为列入一个因子，因此 0.30 是最低的标准。

们按每个国家在两个因子里的分数来作图，就能显示这些国家在二维空间里的位置以及多维数据的结构。一些异常数据也能马上发现，如图 11.37 中的土耳其。

表 11.49　　按国家的因子数值

	Factor 1	Factor 2		Factor 1	Factor 2
Australia	.60240	−.00412	New Zld	.02745	.50339
France	.40991	−.38541	Portugal	−1.29255	.73068
W. Germany	.71182	−1.11358	Spain	−.46100	.45072
Greece	−.87954	.05024	Sweden	.95210	.99570
Iceland	.45184	1.24672	Turkey	−2.39311	.14101
Italy	−.27384	−.52521	UK	.46877	−1.11533
Japan	.21801	−2.22778	USA	1.45773	1.25299

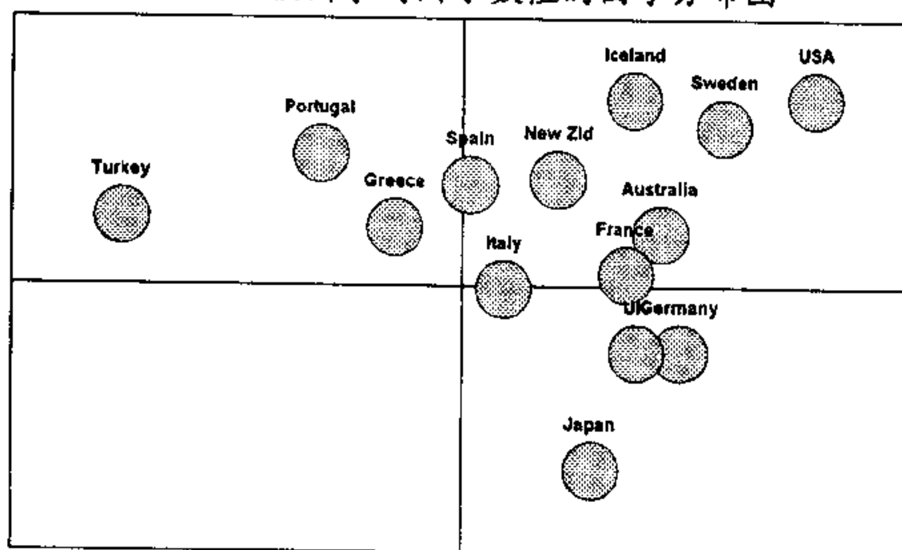


图 11.37　按国家的因子数值的因子分布图

11.7.3.4. 因子分析

因子分析是在主要成分分析的基础上发展起来的。有的心理学家认为用上述方法计算出来的结果总是第一个成分占的比重最大。这可能和我们所定的参照矢量有关，如果我们保持一个矢量的位置不变，而又调整参照矢量，那不是可以减少含混，改善我们对矢量的解释吗？其中的一个简单的调整方法就是旋转参照矢量，这称为因子旋转。在图 11.38 的 (a) 里，OA 是一个矢量，假定我们采取形心法取得一点 A，F_1 和 F_2 分别是第一个因子和第二个因子负荷的值：0.40 和 0.80。如果我们以 O 轴心，把$\angle F_1OF_2$这个直角旋转至$\angle F_1OF_2$ 的位置，而又保持 OA 不变，就可以得到第一个因子和第二个因子负荷的新值：0.75 和 0.50。因子负荷值是改变了，但是它们的相互关系并没有改变，因为

$$0.40^2+0.80^2=0.75^2+0.50^2$$

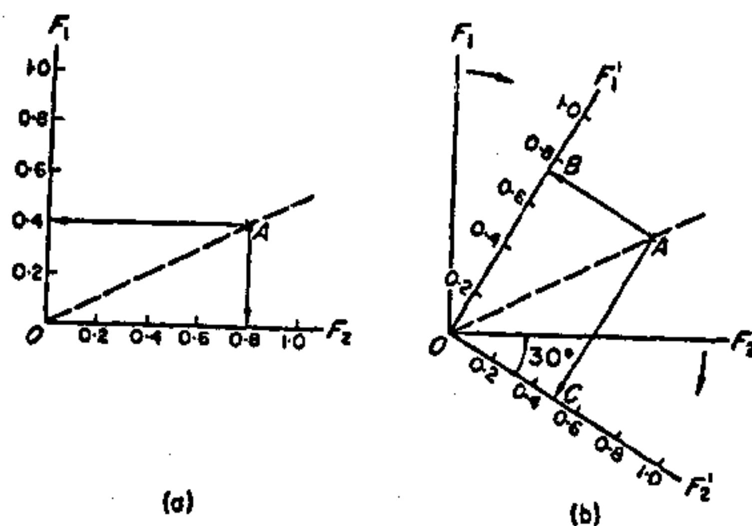


图 11.38

因子旋转的方法较多，较常用的是 VARIMAX。

旋转的结果有时不明显，如图 11.39 所显示的是表 11.47 的 7 个变量的位置，读者可把它和图 11.36 作一比较。但有时变化颇大。

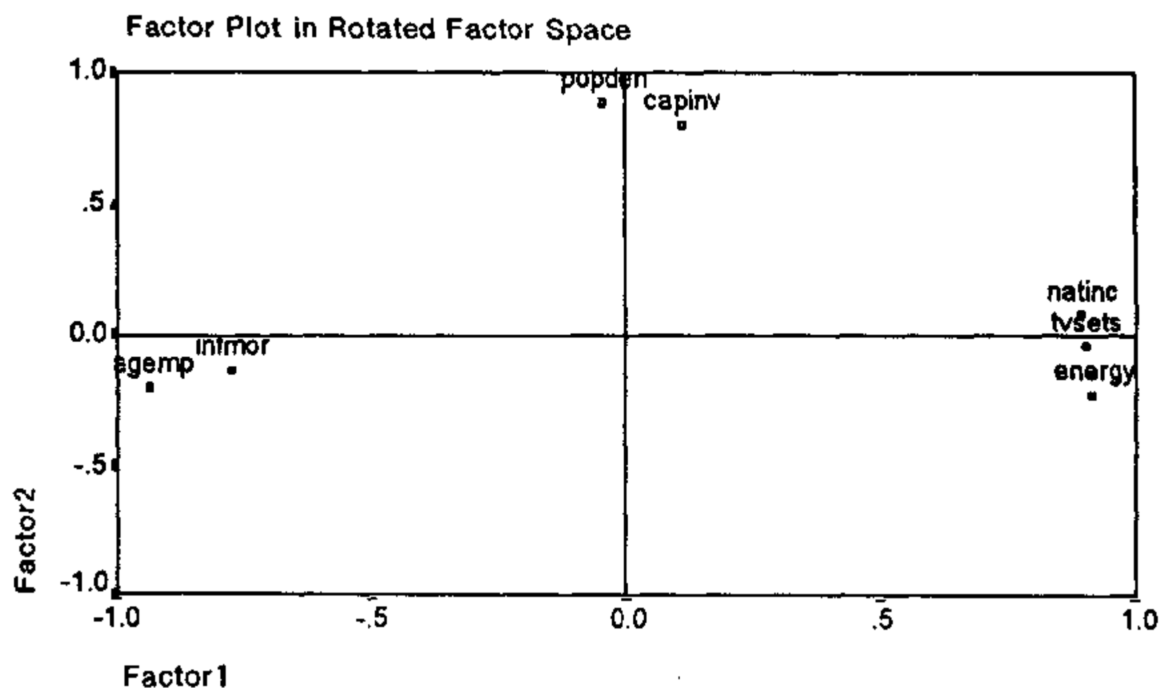


图 11.39　OEDC 经济和社会特征的因子分析图（经过旋转后）

由此可见，使用什么统计方法会影响我们所作的结论。在外语测试界曾围绕外语有没有单一的语言能力（叫做单一能力假设，Unitary Competence Hypothesis，简称 UCH）展开过

热烈的争论。最早提出的是 Oller，他使用主要成分分析法来分析语言测试中的各个部分，得出一个主要因子，称之为 g 因子，g 表示 general（一般的、普遍的）。后来大家对此提出异议，使用因子旋转的因子分析法，对用主要成分分析法所得到的数据重新分析，得出好几个因子。Vollmer (1983) 指出，"主要成分分析法作为检验 UCH 的方法是特别不合适。……因为它倾向于过高估计第一个因子的作用。" Oller (1983) 自己也作过重新分析，最后坦白承认，"关于语言能力中有一个无所不包的普遍因子的想法是错误的。"

为什么不同的统计方法会导致不同的结论呢？Woods (1983) 认为，主要成分分析法是一种描写性技术，它没有什么假设，而仅对数据提供另一种观点。而因子分析却需要研究人员对数据提出一个模型作为假设。不过主要成分分析法作为因子分析的第一步，还是有必要的。

但是不管是用主要成分分析法还是因子分析法，我们也不要轻率地下结论，因为经验告诉我们，从一个总体抽出不同样本来作分析往往会得到不尽相同的结果。这个问题在估算平均值或比例时可以作显著性检验，但使用这些方法时，却不大容易检验。因此我们必须经过多次重复观察，看所得出的因子结构模型是否一致，才能下结论。验证性因子分析 (Confirmatory Factor Analysis) 是一种以验证模型为目标的新的因子分析法由此而产生，我们将在下面 12.7.7.1 节加以介绍。

11.7.4. 聚类分析

聚类分析 (Cluster Analysis) 也是一种常用的多元分析方法，和观察相关程度的多元回归分析与观察差异程度的多元方差分析不同，聚类分析观察的是事物的相近（或相似）程度。既然"物以类似，人以群分"，我们就有必要观察事物之间有些什么共享的或差异的特征，并根据这些特征进行分类。聚类分析的根本目的是寻找同质性组别。在生物学里，我们可以用聚类分析来对动物和植物分类；在医学里，我们用它来寻找病因和病情发展阶段；在市场调查，我们用它来找出具有相同购买习惯的顾客；在社会语言学里，我们用它来区分不同类型的说话人的群体；在语言习得里，我们用它来了解儿童的语言发展阶段。

11.7.4.1. 距离矩阵

聚类分析的基本出发点是考察事物相近性，相近性表示为事物之间的距离的远近。距离近，相似程度就大；距离远，相似程度就小。我们仍以表 11.47 的数据为例，假定我们想了解这些国家在第三个变量（国民收入）和第六个变量（能源）方面的距离，我们可以以国民收入作为 X 轴，以能量作为 Y 轴，作出图 11.40：

如果我们想具体了解两个国家，如美国和意大利的距离，我们必须找出在两个国家这两个变量之间的差，然后根据勾股定理来找出其距离：

聚类分析正是根据距离矩阵来对事物相互之间的关系作出分析。计算距离矩阵的方式有多种，Woods (1986) 进一步说明，距离（非相似性）和相关的关系可表示为 $1-r^2$，这就是说，如果两个学生的分数完全是一样的，非相似性就是 0。在一个相关距阵里，我们先找出相关最高的变量，把它们归为一组，然后在找次相关的变量，又把它们分为另一组，……就可以聚成不同的类别。

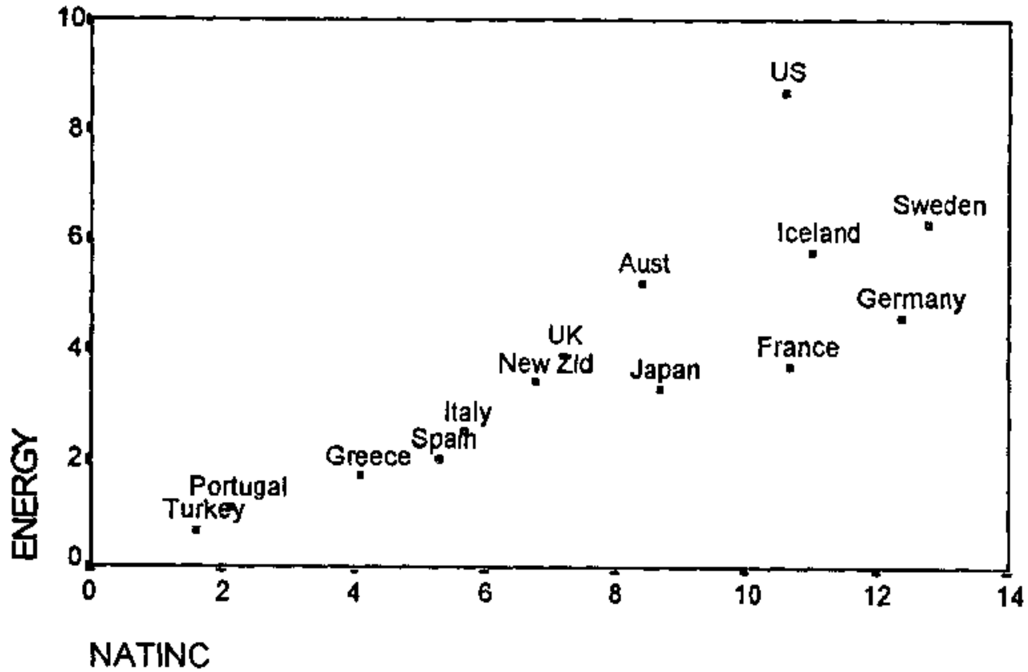


图 11.40 按照国民收入和能源来测量一些国家的距离

表 11.50

美国和意大利在国民收入和能源的距离

	X_3	X_6
美国	10.6	8.7
意大利	5.7	2.5
差	4.9	6.2
d=	$\sqrt{4.9^2+6.2^2}=7.9$	

计算美国和意大利的距离的示意图

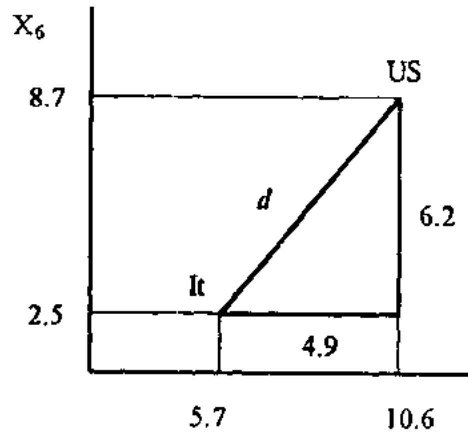


图 11.41 计算美国和意大利的距离的示意图

11.7.4.2. 分层聚类分析

分层聚类分析 (Hierarchical Cluster Analysis) 是使用最广泛的一种聚类分析。我们仍以表 11.47 的数据为例，用 STATISTICA 中的 Joining (tree clustering) 的程序来生成出下列结果，以兹说明。SPSS 所得的结果也是一样的，但 STATISTICA 可以分页给出结果，便于说明。

首先，程序可以给出距离矩阵，如表 11.51：

表 11.51　　OECD 的七个经济和社会因素的距离矩阵

	POPDEN	AGEMP	NATINC	CAPINV	INFMOR	ENERGY	TVSETS
POPDEN	0	498	510	507	501	522	468
AGEMP	498	0	74	65	95	78	111
NATINC	510	74	0	13	158	17	85
CAPINV	507	65	13	0	149	19	87
BINFMOR	501	95	158	149	0	160	

由此可见变量之间的距离是很不一样的，最远的可达 522（人口密度和能源），最近的只有 13（国民收入与工业投入）。根据这个距离矩阵，我们可以先算出合并方案（Amalgamation Schedule），如表 11.52：

表 11.52

Amalgamation Schedule (fact. sta)

Single Linkage

Euclidean distances

	Obj. No. 1	Obj. No. 2	Obj. No. 3	Obj. No. 4	Obj. No. 5	Obj. No. 6	Obj. No. 7
13.23367	NATINC	CAPINV	0	0	0	0	0
16.53148	NATINC	CAPINV	ENERGY	0	0	0	0
64.95975	AGEMP	NATINC	CAPINV	ENERGY	0	0	0
84.59752	AGEMP	NATINC	CAPINV	ENERGY	TVSETS	0	0
94.95789	AGEMP	NATINC	CAPINV	ENERGY	TVSETS	INFMOR	0
468.2040	POPDEN	AGEMP	NATINC	CAPINV	ENERGY	TVSETS	INFMOR

这个表的意思是第一步根据 12.23367 的距离，把 NATINC 和 CAPINV 列为一类，第二步根据 16.53148 的距离，再把 NATINC，CAPINC 和 ENERGY 一类，以此类推，一直到第六步。这也就是说，这些变量可以按 6 个层次来归类，请注意归类的距离标准是不一样的，可表示为下图：

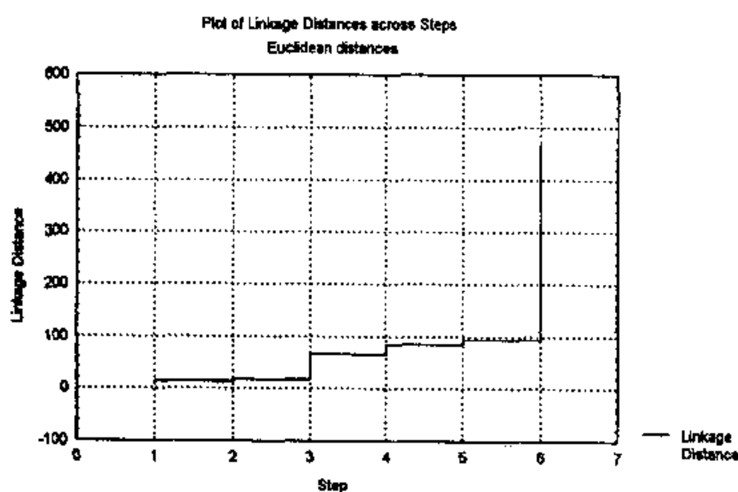


图 11.42

最后为了提高目击的效果，我们可以把这些层次关系表示为一个树形结构，如图11.43：

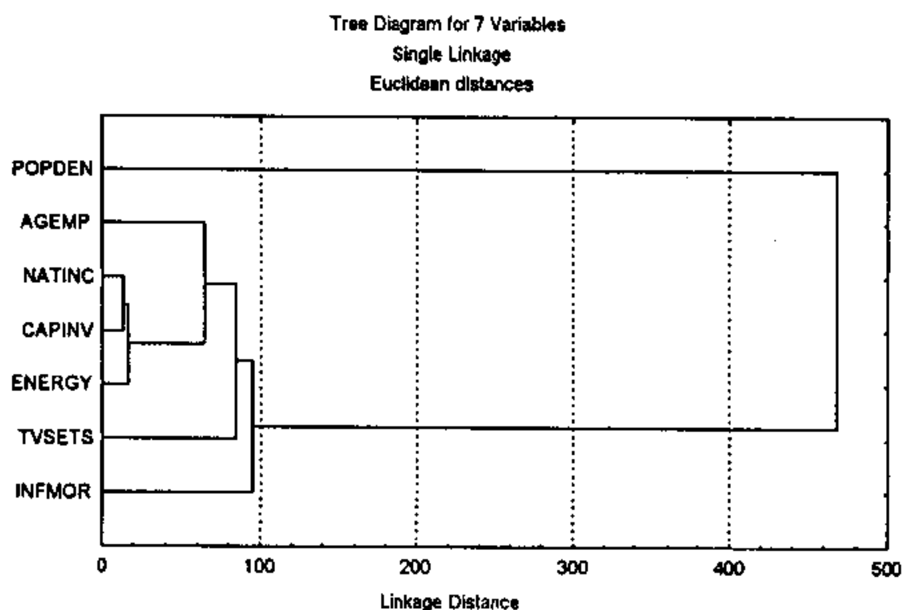


图 11.43

读者不妨把分层聚类分析的结果和因子分析的结果加以比较，以见这两种分析手段的异同。它们都以变量之间的关系为基础，但一种分析的目的是简约因子，另一种分析目的却是观察变量能否根据其接近程度而组成分层的类别。这两种方法可以互为补充。

我们也可用个案，而不是变量来作分层聚类分析，了解各个国家之间的归类，如图11.44：

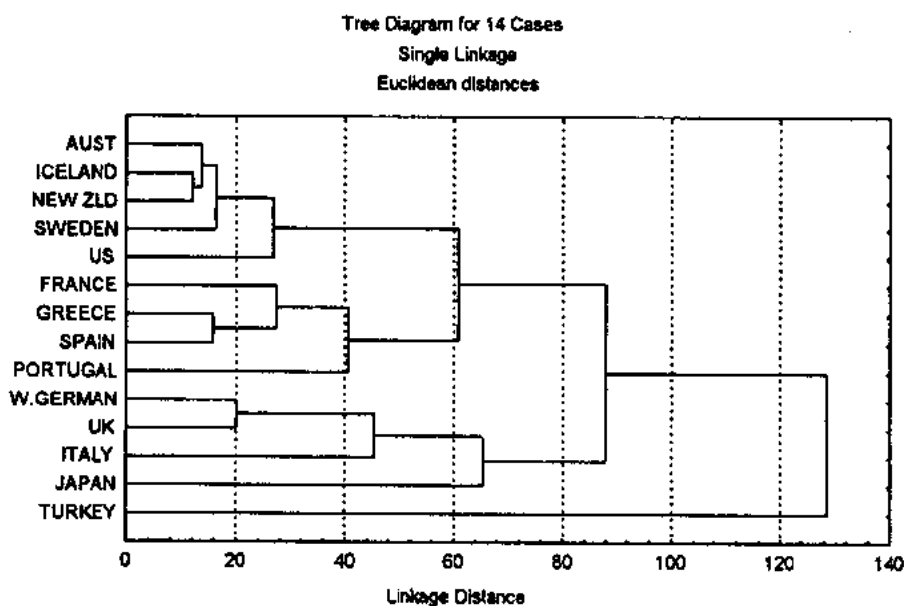


图 11.44

11.7.5. 多维度量表

多维度量表(Multidimensional Scaling)也属于一种聚类分析，它的出发点也是观察事物的距离或非相似性。但是多维度量表的目的并非象聚类分析那样把实验因素归类，而是根据距离矩阵来建立一个表示因素之间关系的示意图。所以多维度量表所要解决的问题和因子分析的相似，不过它采用了图示的方法。仍以表 12.47 的数据为例，在 STATISTICA 的 Multidimensional Scaling 可先计算出各个变量在两个维度上的值，然后再作出一个两维图，图 11.45：

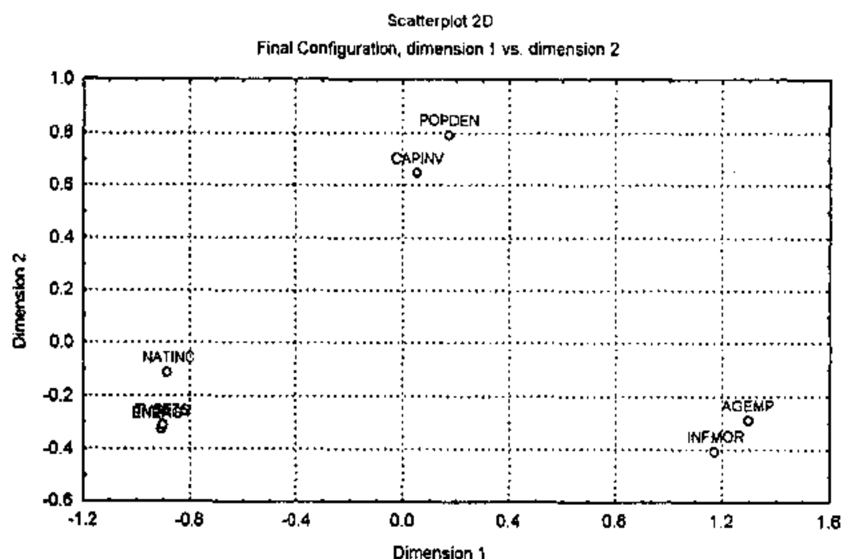


图 11.45 OEDC 的七个因素的多维度表

从图中看，变量是组合成三个群体的。如果我们要求程序拟合成三维，也可以作出一个三维图。图 11.46 是用 STATISTICA 作的，图 11.47 是用 Harvard ChartXL 作的另一种三维面积图。

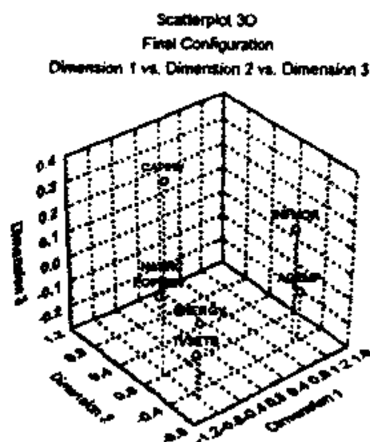


图 11.46 OEDC 的七个因素的三维图

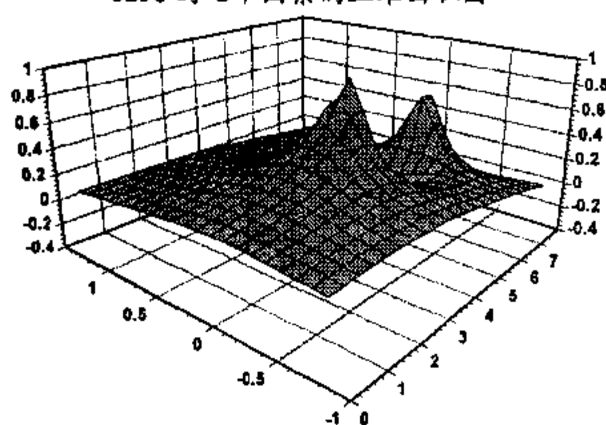


图 11.47 OEDC 的七个因素的三维面积图

表 11.53

Starting Configuration (multi. sta)

(last final)

	DIM. 1	DIM. 2
POPDEN	.141926	.815017
AGEMP	1.349408	-.225579
NATINC	-.858178	.006436
CAPINV	-.045576	.521143
INFMOR	1.195167	-.372749
ENERGY	-.895395	-.381640
TVSETS	-.887351	-.362628

多维度量表在语言研究中应用得也很广泛，例如 Miller & Nicely (1955) 用它来研究 16 个英语辅音的感知，他们观察了声音歪曲和噪音对 5 个发音特征（维度）的影响：轻浊音、鼻音、塞擦音、发音的时间和部位。他们发现这些特征的感知是相对独立的，而轻浊音和鼻音最不受各种实验条件的影响。他们使用这两个维度来作图，可见浊鼻音是属于一类的，浊塞音和擦音属于另一类，而轻塞音和擦音又是另一类。又有一些研究语义的人用多维度量表来研究语义空间，例如 Rips 等人发现人们是使用两个维度（凶猛性和体积）来表示动物的。

11.7.6. 判别分析

判别分析（Discriminant Analysis）和聚类分析刚好相反。聚类分析是根据一些特征来对变量分类，而判别分析是根据一些特征来判别变量应否属于一类。例如我们把世界上的国家分为发展国家和发展中国家，而又收集了这些国家的社会和经济变量，就可以根据这些参数来判别在发展的（或发展中的）国家群里，有哪些是归类不当的。一个大学里有不同年级的学生，如果我们对这些学生从不同的角度来进行测量，也可以根据测量的分数来判别在这些年级里有哪些学生是不属于他们所在的年级的。要判断考古发掘中的文物是否属于同一时期的，我们也使用这种方法。例如，一些考古学家认为发掘出来的希腊的瓷器碎片可分为两类：一类是雅典时代（Attic）的，另一类是厄立特里亚时代的（Eritrean），然后对这些碎片的六种化学金属氧化成分进行分析，就可根据化学分析结果来判别。为了说明的方便，我们只取两种分析：镁（Mg）和钙（Ca）：

如果我们用 MG 和 CA 来作一个散布图（scatter plot），就可见属于雅典时代的集中在右上部分，而属于厄立特里亚时代的在右下部分，从中可以作一条线，这条线就是线性判别边

界线（linear discriminant boundary），而它的方程就是 3.24MG＋1.35CA－5.01＝0。这个方程的左边就是判别函数（discriminant function）。这个方程是怎样计算出来的，后面还要交代。

表 11.55　　不同时代的希腊瓷器碎片的镁（Mg）和钙（Ca）成分

Attic				Eritrean			
编号	Mg	Ca	d	编号	Mg	Ca	d
1	4.9	5	4.1	1	2.5	5.8	－4.7
2	4.4	4.8	2.8	2	2.3	4.9	－4.2
3	6.4	11.2	0.6	3	5.5	8.7	1.1
4	4.9	6.6	1.9	4	2.4	7.0	－6.7
5	5.5	5.4	5.5	5	2.5	5.7	－4.6
6	4.9	5.4	3.6	6	2.5	4.7	－3.3
7	6.3	5.1	8.5	7	2.7	4.2	－1.9
8	7.5	8.5	7.8	8	2.6	4.9	－3.2
9	2.9	5.2	－2.6	9	2.0	5.6	－6.1
10	4.6	4.6	3.7	10	5.1	7.7	1.1
11	5.3	4.4	6.2	11	4.4	12.8	－8.0
12	4.4	4.4	3.3	12	4.2	9.4	－4.1
13	4.4	3.9	4.0	13	2.7	5.6	－3.8
				14	3.3	7.7	－4.7
平均值	5.1	5.7	3.8		3.2	6.7	－3.8

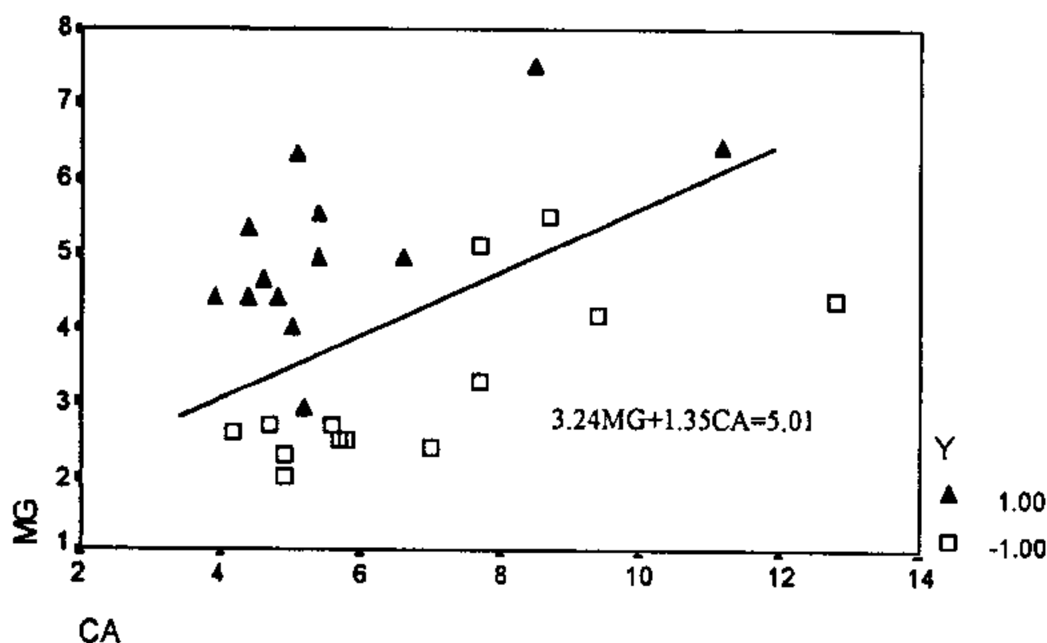


图 11.48　Mg 和 Ca 的散布图

根据这个方程式就可算出判别分数，d 值（discriminant score)，如雅典组的第一号为：

(3.24)(4.9) − (1.35)(5.0) −5.01=4.1

而厄立特里亚组的第一号为：

(3.24)(2.5) − (1.35)(2.8) −5.01=−4.7

以此类推，就可以计算出两组的 d 值出来。判别函数和判别分数在判别分析中至关重要，因为它们有两个代数特征：

(1) 在判别边界线上任何一点的判别分数都是 0，因为它满足了判别边界方程的要求。

(2)在判别边界线一边的判别分数点的符号和另一边的符号刚好相反，因此一为正值，另一为负值。从表 11.54 可见，雅典时代的碎片为正值，而厄立特里亚时代的为负值。

表 11.55

Classification function coefficients

(Fisher's linear discriminant functions)

Y	=	Eritrean	Attic
Ca		.996	−.356
Mg		1.17	4.41
(Constant)		−5.93	−10.94

如果在一组应为正值（或负值）的数值中有相反的数值，这个数值就不应属于这个组，例如雅典时代的第 9 号，厄立特里亚时代的第 3 号和第 10 号。换句话说，雅典时代第 9 号应属厄立特里亚时代，而厄立特里亚时代的第 3 号和第 10 号应属雅典时代。

让我们再用 SPSS（从主页去 Classify/Discriminant）来处理同样的数据，看所得到的结果。首先是两组的判别函数，如表 11.55。如果我们求出 Attic 和 Eritrean 两组之差，求可以得到方程 3.24Mg−1.35Ca−5.01。其次，程序产生一个具体的个案表，说明哪一些个案是不属于所属的组别，如表 11.56 中标有两个星号的都应该属于另一组的。表中还提供判别分数的两个概率：一个是条件概率，表示为 P(D|G)，另一个是后验概率(posterior probability)，表示为 P(G|D)。例如第 1 号个案原属雅典时代组，其条件概率为 0.9093，其后验概率为 0.9840，其不属于这个组的后验概率为 0.0160。我们主要的应看后验概率，第二个后验概率是 1−第一个后验概率，故 1−0.9840=0.0160。其判别分数为 1.5440。SPSS 所给的这个判别分数和表 11.54 的 d 有所不同，它所给的是经过转换的分数，目的是保证整体的平均判别分数等于 0。但是这个转换分仍然有正负之分，再加上前面的星号，意义仍是十分清楚的。

表 11.56

Case	Actual	Highest	Probability		2nd Highest		Discrim
Number	Group	Group	P (D/G)	P (G/D)	Group	P (G/D)	Scores
1	1	1	.9093	.9840	−1	.0160	1.5440
2	1	1	.7069	.9408	−1	.0592	1.0541
3	1	1	.2457	.6461	−1	.3539	.2693
4	1	1	.5027	.8760	−1	.1240	.7599
5	1	1	.5330	.9960	−1	.0040	2.05
35	1	1	.9346	.9728	−1	.0272	1.3480
7	1	1	.0871	.9998	−1	.0002	3.14

续前表

Case Number	Actual Group	Highest Group	Probability P (D/G)	P (G/D)	2nd Highest Group	P (G/D)	Discrim Scores
12	1	1	.1454	.9996	−1	.0004	2.8859
9	1**	−1	.6728	.9333	1	.0667	−.9057
10	1	1	.9659	.9755	−1	.0245	1.3873
11	1	1	.3798	.9980	−1	.0020	2.3084
12	1	1	.8572	.9647	−1	.0353	1.2502
13	1	1	.9481	.9817	−1	.0183	1.4952
14	−1	−1	.7323	.9914	1	.0086	−1.6701
15	−1	−1	.8917	.9849	1	.0151	−1.4642
16	−1**	1	.3203	.7431	−1	.2569	.4362
17	−1	−1	.2947	.9988	1	.0012	−2.3758
18	−1	−1	.7694	.9902	1	.0098	−1.6211
19	−1	−1	.8438	.9630	1	.0370	−1.1310
20	−1	−1	.4983	.8739	1	.1261	−.6508
21	−1	−1	.8286	.9611	1	.0389	−1.1114
22	−1	−1	.4054	.9978	1	.0022	−2.1600
23	−1**	1	.3300	.7534	−1	.2466	.4560
24	−1	−1	.1239	.9997	1	.0003	−2.8666
25	−1	−1	.9144	.9837	1	.0163	−1.4354
26	−1	−1	.9929	.9787	1	.0213	−1.3369
27	−1	−1	.7394	.9912	1	.0088	−1.6606

SPSS 还可以生成一个区域图，进一步说明这两组数据中的个案分布情况，如图 11.49。图中清楚地标明在 1 组有一个个案应属于 2 组，而在 2 组中有两个个案应属于 1 组。

```
                    All-groups Stacked Histogram

                  Canonical Discriminant Function 1

         4 +                                                         +
           |                                                         |
F          |                                                         |
r        3 +                  11             2                       +
e          |                  11             2                       |
q          |                  11             2                       |
u          |                  11             2                       |
e        2 +                  111      1     2                       +
n          |                  111      1     2                       |
c          |                  111      1     2                       |
y          |                  111      1     2                       |
         1 +             1 11 1112 1   21 22222 2 2 2 2              +
           |             1 11 1112 1   21 22222 2 2 2 2              |
           |             1 11 1112 1   21 22222 2 2 2 2              |
           |             1 11 1112 1   21 22222 2 2 2 2              |
           X---------+---------+---------+---------+---------+-------X
          out      -4.0      -2.0        .0       2.0       4.0     out
   Class   1111111111111111111111111111111112222222222222222222222222222222
Centroids                     1             2
```

图 11.49

我们在这里仅用两个变量和两个组来说明判别分析的基本原理，实际上在多元分析里，往往牵涉到多个变量，多个组别，使用这些现成的软件照样可以得出令人满意的结果。

11.7.7. Lisrel 模型

Lisrel (Linear Structural Equation Model 的简称）模型是 70 年代瑞典统计学家 Jöreskog 和 Sörbom 提出来的一种企图把经济测量学和心理测量学的特征结合在一起的多元分析方法。他们编制了 Lisrel 程序，在美国由 Scientific Software 公司发行，经过不断更新，已经出版了八个版本。SPSS 和 STATISTICA 都在它们的更新版中增加了这种方法，但是 SPSS (6.0) 尚不能在菜单里进行选择，而必须自行编写命令，STATISTICA (5.0) 则可以通过菜单来运算，比较方便。倒是 Lisrel 程序本身较难操作，需要懂得一些 Fortran 算法语言的知识。Lisrel 模型也有人称为结构方程模型 (Structural Equation Model) 或路径分析 (Path Analysis)，但实际上，它包括两大部分：

1. 测量模型，描写观察到的变量的信度和效度，说明观察的变量和潜在变量 (latent variables) 的关系，潜在变量是怎样依赖于观察变量，怎样由观察变量来表示的。验证性因子分析 (Confirmatory Factor Analysis) 就是其中的一种方法。

y 的测量模型表示为：

$$y=\Lambda_y\eta+\varepsilon \tag{11.43}$$

x 的测量模型可表示为

$$x=\Lambda_x\xi+\delta \tag{11.44}$$

2. 结构方程模型，描写潜在变量的因果关系，给能解释的或不能解释的方差赋值。

所以 Lisrel 模型可用来估算一些线性结构方程的未知的系数，特别适合于用来解释包括有潜在变量、自变量和依变量都有误差、相互影响、共时、互相依存等因素的模型。结构方程模型表示为：

$$\eta=B\eta+\Gamma\xi+\zeta \tag{11.45}$$

Lisrel 使用希腊文字母来作标记，因为电脑程序没有希腊文，我们只好用拉丁文来代替，现将常用的列表如下：另外还有一些路径的表示方法：(1) 观察变量 x，y 用方框表示；(2) 潜在变量 ξ，η 用圆框表示；(3) 误差变量 ζ，δ，ε 不加框；(4) 两个变量之间的单向直线箭头表示一个变量直接影响另一个变量；(5) 弧形双向箭头表示两个变量可能相关，但不一定有直接关系；(6) 在潜在变量中，ξ 为自变量，不应有单向箭头指向它；η 为依变量，所有指向 η 的箭头来自 ξ 变量和 η 变量。

表 11.57　Lisrel 所使用的符号的含义

含义	符号
观察变量	x，y
潜在变量	ξ (ksi，xi)，η (eta)
误差变量	ζ (zeta)，δ (delta)，ε (epsilon)
因子负荷	Λ_y，Λ_x (lambda)
结构参数	B (beta)，Γ (gamma)
协方差矩阵	Φ (phi)，Ψ (psi)
误差协方差矩阵	Θ_ε，Θ_δ (theta)
Λ_x 的估算	$\hat{\Lambda}_x$

11.7.7.1. 验证性因子分析

验证性因子分析和一般的（可称为探索性的）因子分析不同：在探索性因子分析里，我们观察实验数据，以发现它们的特征和有意义的关系．对数据并不强加什么模型，所以这种分析可以生成结构、生成模型或生成假设。在验证性因子分析里，我们先建立一个模型，用相对少的参数来解释实验数据。这个模型建筑在事先已知的信息的基础上，以一种假设或理论的形式出现。探索性因子分析常用来发现和评估潜在的方差和共差源，在实验的初期是很有用的，但是随着我们对社会和心理测量的性质的了解日益增加，光依赖探索性因子分析，就显得很不够。

让我们看一个具体的例子，Holzinger & Swineford (1939) 曾对芝加哥的 145 名中学生作了 26 个心理测验，Jöreskog 和 Lawley 选了其中 9 个测验的数据，其中有三个和视觉有关的：视觉显示 (VIS-PERC，x1)、立方形 (CUBES，x2)、菱形 (LOZENGES，x3)；有三个是和语言能力有关的：段落写作 (PAR-COM，x4)、句子写作 (SEN-COM，x5)、词义 (WORD-MEAN，x6)；有三个和速度有关：加法 (ADDITION，x7)、数点 (COUNTDOT，x8)、直一弧大写字母 (S-C-CAPS，x9)。他们假定视觉、语言和速度都是潜在变量 (用圆框表示，标号为 ksi1，ksi2 和 ksi3)，而它们互有联系 (用弧形双箭头表示，标号为 phi31，phi21，phi32)。每一个变量和它所包括的三个成分都有关的 (用直线单箭头表示，标号为 lambda11，lambda21，lambda31…lambda93)。9 个变量都有误差 (用直线单箭头表示，标号为的 delta1，delta2，delta 3…delta 9)。这可以表示为一个路径图：

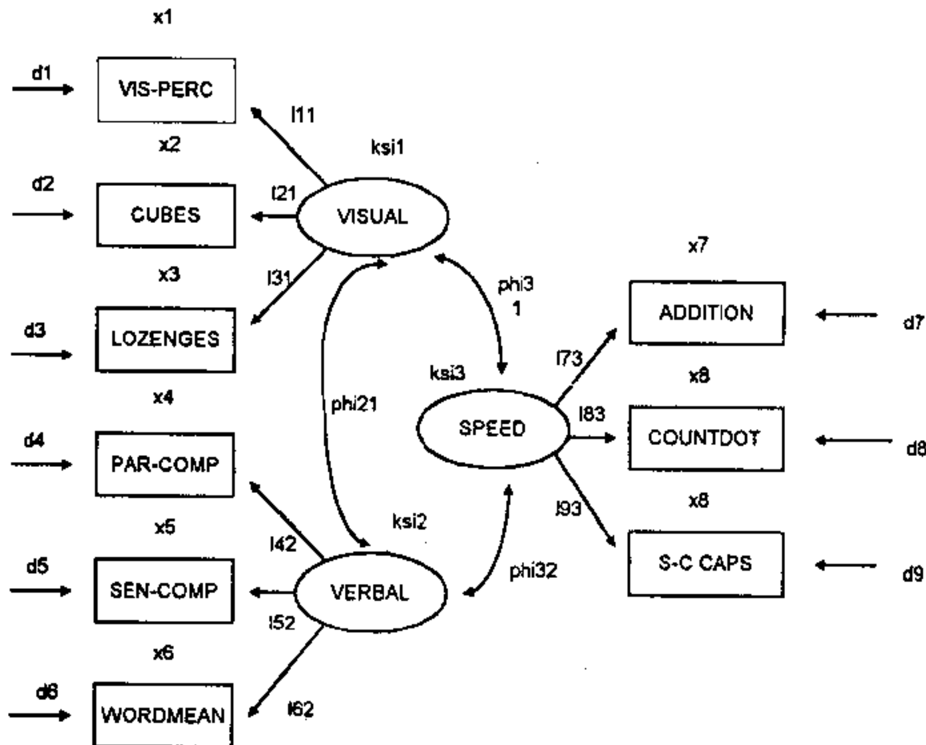


图 11.50 九个变量的相互关系的路径示意图

验证性因子分析就是要验证所提出的这个模型。和一般的因子分析一样，验证性因子分析的基础是相关矩阵或协方差矩阵。在STATISTICA里，我们要先把矩阵作为矩阵格式的文件输入，除了变量名外，还必须增加每个变量的MEANS（平均值），STD. DEV.（标准差），NO.CASES（个案数），MATRIX（矩阵），前两者可以省略，MATRIX需填进代号，1为相关矩阵，2为相似形矩阵，3为非相似性矩阵，4为协方差矩阵。如表11.58：

表11.58　　STATISTICA输入矩阵的格式

VIS－PERC	1.00								
CUBES	0.318	1.00							
LOZENGES	0.436	0.419	1.00						
PAR-COM	0.335	0.234	0.232	1.00					
SEN-COMP	0.304	0.157	0.283	0.722	1.00				
WORDMEAN	0.326	0.195	0.350	0.714	0.685	1.00			
ADDITION	0.116	0.057	0.056	0.203	0.246	0.170	1.00		
COUNTDOT	0.314	0.145	0.229	0.095	0.181	0.113	0.585	1.00	
S-C-CAPS	0.489	0.239	0.361	0.309	0.345	0.280	0.408	0.512	1.00
MEANS									
STD. DEV.									
NO. CASES	154								
MATRIX	1								

然后选择SEPATH，在页面上有两种选择，一是Edit with Path Tools，一是Use Path Wizards。前者是供熟悉验证因子分析命令的人使用的，后者较为简单，只要作出选择即可。选了Use Path Wizards后，应选Confirmatory Factor Analysis。进入该页面后，可以按照我们的模型键入潜在变量（如Visual，Verbal，Speed等），并在Manifest栏目中选择它所包括的显露性变量，即9个心理测验中的几个，如键入Visual后选VIS－PERC，CUBES，LOZENGENS等等。但是在Use Path Wizards里，不能规定潜在变量之间的关系，因此还要到Edit with Path Tools里，规定Visual，Verbal，Speed三者的相互联系的关系，因为它们是双箭头的关系，故应在页面上选X←→Y（单箭头的选X ──→Y）。最后在OK（Execute Current Model）按键就能运算，其结果如11.59。这个表给出每条路径的参数值、误差、t值和显著性意义。一般的参数值应为正值，如果是相关系数，就不能超过1。误差是表示参数估算的精确性的，误差小就精确度高，误差大就精确度低。但是什么算大，什么算小，也很难界定。比较容易判断的t值，它不受测量单位的影响。t值是参数值和误差的比率，在t值后面的是p值，表示显著意义。

表 11.59

Model Estimates (lawley. sta)

	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(speed) -19- (verbal)	.321	.093	3.452	.001
(speed) -20- (visual)	.512	.096	5.306	.000
(verbal) -26- (visual)	.543	.086	6.327	.000
(visual) -1-> [VIS-PERC]	.673	.091	7.392	.000
(visual) -2-> [CUBES]	.513	.092	5.550	.000
(visual) -3-> [LOZENGES]	.684	.091	7.514	.000
(verbal) -4-> [PAR-COMP]	.867	.070	12.348	.000
(verbal) -5-> [SEN-COMP]	.830	.072	11.609	.000
(verbal) -6-> [WRD-MNG]	.826	.072	11.525	.000
(speed) -7-> [ADDITION]	.661	.085	7.762	.000
(speed) -8-> [CNT-DOT]	.797	.084	9.430	.000
(speed) -9-> [S-C-CAPS]	.681	.085	8.009	.000
(DELTA1) --> [VIS-PERC]	....	....	....	....
(DELTA2) --> [CUBES]	....	....	....	....
(DELTA3) --> [LOZENGES]	....	....	....	....
(DELTA4) --> [PAR-COMP]	....	....	....	....
(DELTA5) --> [SEN-COMP]	....	....	....	....
(DELTA6) --> [WRD-MNG]	....	....	....	....
(DELTA7) --> [ADDITION]	....	....	....	....
(DELTA8) --> [CNT-DOT]	....	....	....	....
(DELTA9) --> [S-C-CAPS]	....	....	....	....
(DELTA1) -10- (DELTA1)	.548	.097	5.642	.000
(DELTA2) -11- (DELTA2)	.737	.101	7.301	.000
(DELTA3) -12- (DELTA3)	.532	.097	5.467	.000
(DELTA4) -13- (DELTA4)	.248	.051	4.820	.000
(DELTA5) -14- (DELTA5)	.311	.054	5.786	.000
(DELTA6) -15- (DELTA6)	.318	.054	5.883	.000
(DELTA7) -16- (DELTA7)	.563	.087	6.449	.000
(DELTA8) -17- (DELTA8)	.365	.089	4.098	.000
(DELTA9) -18- (DELTA9)	.536	.087	6.18	2.000

由此可见，视觉、语言和速度这三者的联系有显著意义，而它们和各自的三个部分的关系也有显著意义，因此我们所提出的模型得到证实。除了观察这些路径的显著意义外，我们还要观察整个模型的拟合程度，一般是看下面几个值：

(1) 卡方 (χ^2) 检验：一般来说，卡方值大表示拟合不太好，卡方值小表示模型拟合得较好。现计算出来的卡方值为 52.618。

(2) 拟合优度指数 (Goodness-of-Fit Indices，GDI) 和调整拟合优度 (AGDI)：指数从 0 到 1，从理论上说，也可以有负值。如果出现负值就说明模型的拟合比任何模型都要差。现

在 GDI 为 0.928，AGDI 为 0.866。

(3) 均方根残余值 (Root Mean Squared Residual，RMR)：它测量拟合残余值的平均值，必须结合方差和协方差来进行观察。如果观察的变量经过标准化处理，这个统计量更能说明问题。均方根残余值可用来比较以相同数据为基础的两个模型的拟合程度。现在标准化的 RMR 为 0.075。

另外 STATISTICA 还提供一些别的统计量，如差异函数 (discrepancy function)，这是使用了某一种参数估算方法后所得到一个函数，越低越好，现为 0.365。

从总体来看，52.618 的卡方值比较大，说明模型的拟合尚不理想。卡方值的大小和样本数以及样本的正态性有关，因此卡方值大说明存在误差。要进一步拟合模型就必须依靠一些专业性的软件，Lisrel 有一个修正指数的功能，它能告诉我们有哪些数值需要修正。在表 11.60 里可见最大的修正指数是 24.74，是 P-C-CAPS 和视觉之间的 lambda X，应予释放。原来模型只在速度和 P-C-CAPS 之间建立路径，可能不全面，因为大写字母的辨认有速度的问题，但也有视觉的问题，它牵涉到一些直线和弧线的视觉判断。Lisrel 可以加一个语句来释放 P-C-CAPS，这样卡方值就能降低至 28.86，拟合优度提高到 0.958，标准化 RMR 为 0.045。原来视觉和 P-C-CAPS 没有路径，而现在增加参数值 0.459，而视觉和它的路径则从 0.677 降至 0.412。

表 11.60

Modification Indices and Estimated Change

Modification Indices for lambda X

	Visual	Verbal	Speed
VIS-PERC	.000	.268	3.918
CUBES	.000	.665	.969
PAR-COMP	.000	.032	1.346
SEN-COMP	.342	.000	2.059
WORDMEAN	.277	.000	.303
ADDITION	10.502	.178	.000
COUNTDOT	2.738	10.078	.000
S-C-CAPS	24.736	10.040	.000

11.7.7.2. 结构方程模型

结构方程模型是一种应用范围很广的多元统计方法，它既可以用来计算多元回归和多元方差分析，也可以用来处理非正态数据和序数变量。它既可以考察直接观察到的变量的因果关系，也可以检验具有潜在变量的模型。可观察的变量叫做显露 (manifest) 变量，和潜在 (latent) 变量相对应，它们还可以分为外在的 (exogenous) 和内在的 (endogenous) 变量，因此在一个模型里就可以有显露外在变量、显露内在变量、潜在外在变量和潜在内在变量，这四种变量形成错综复杂的关系。

结构方程模型的应用范围甚广，Jöreskog & Söbom (1989) 举了很多典型的例子，这里限于篇幅，无法多作介绍。在 10.7.2.4 里我们谈到一种多维度多方法设计，现在就举一个这样的例子来做说明。Jaccard 等人在 1975 年用四种方法来调查人们对抽烟态度和对处死刑的态度：(1) 语义微分；(2) Likert 量表；(3) Thurstone 量表；(4) Guilford 量表。Kenney (1979) 根据他们的数据整理成一个相关矩阵：

表 11.61 四种方法调查两种态度的相关矩阵

C1	1.00							
C2	.780	1.00						
C3	.810	.770	1.00					
C4	.760	.710	.810	1.000				
P1	.290	.230	.190	.100	1.000			
P2	.280	.290	.180	.090	.840	1.000		
P3	.260	.310	.240	.080	.810	.890	1.000	
P4	.270	.240	.230	.150	.840	.910	.850	1.000
Means								
Std.Dev.								
No.Cases	35							
Matrix	1.000							

然后在提出一个模型，对吸烟态度的（C）的调查有4种方法（C1到C4），对死刑态度（P）的调查也有4种方法（P1到P4），我们不但要了解C和C1，C2，C3，C4，P和P1，P2，P3，P4的关系，而且还想了解这4种方法在调查两种态度时的关系，P和C的关系，如图11.51。

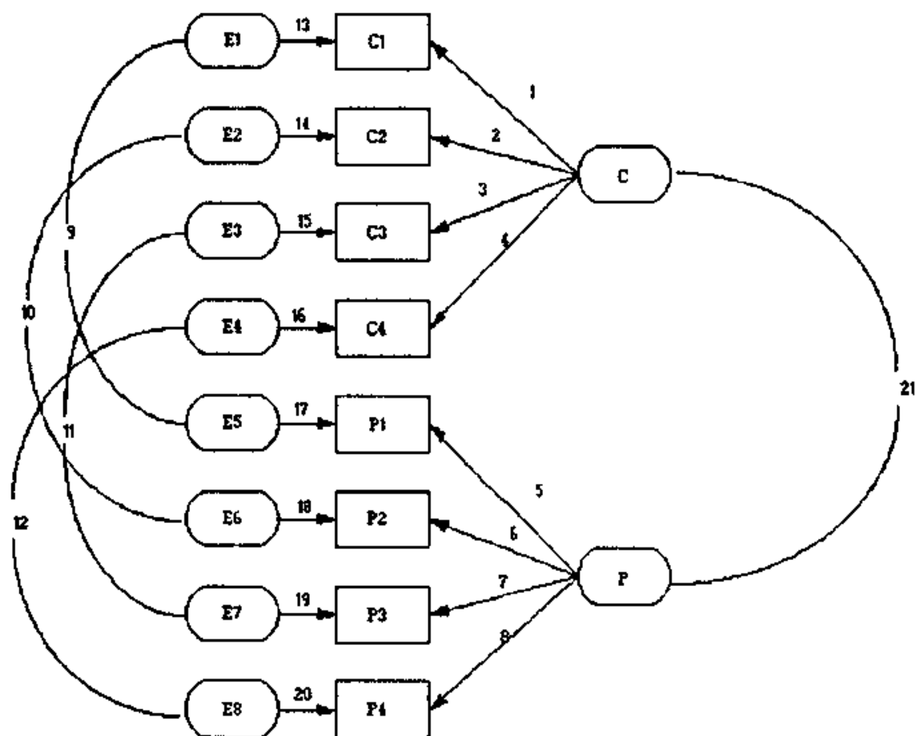


图 11.51 使用四种方法调查两种态度的相互关系的路径示意图

用 STATISTICA 所得的总体的结果如表 11.62。

表 11.62

Basic Summary Statistics (jaccard.sta)	Value
Discrepancy Function	.305
ML Chi-Square	10.354
Degrees of Freedom	15.000
p-level	.797
RMS Standardized Residual	.049
GFI	.930
AGFI	.831

从整个模型的拟合来看，结果是令人满意的，但是样本数只有 35，这是我们感到不放心的。再看具体路径的显著意义，C 和 C1，C2，C3，C4，P 和 P1，P2，P3，P4 这 8 条路径都有显著意义，说明对每一种态度调查而言，这 4 种方法都是有效的。但是 4 种方法之间，E1 和 E5，E2 和 E6，E3 和 E7，E4 和 E8，却没有显著意义。这是因为样本太少，难以验证方法。同样的，对吸烟和对死刑的态度，是两种无甚联系的态度，因此也没有显著意义。

表 11.63

Model Estimates (lawley.sta)

	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(C) -1-> [C1]	.893	.135	6.619	.000
-2-> [C2]	.852	.139	6.121	.000
-3-> [C3]	.914	.132	6.904	.000
-4-> [C4]	.860	.137	6.280	.000
(P) -5-> [P1]	.867	.134	6.490	.000
-6-> [P2]	.957	.125	7.674	.000
-7-> [P3]	.916	.132	6.956	.000
-8-> [P4]	.964	.130	7.422	.000
(E1) -13-> [C1]	.449	.076	5.880	.000
(E2) -14-> [C2]	.532	.079	6.745	.000
(E3) -15-> [C3]	.400	.078	5.146	.000
(E4) -16-> [C4]	.506	.077	6.548	.000
(E5) -17-> [P1]	.477	.065	7.320	.000
(E6) -18-> [P2]	.263	.060	4.364	.000
(E7) -19-> [P3]	.409	.060	6.808	.000
(E8) -20-> [P4]	.329	.058	5.638	.000
(E1) -9- (E5)	.183	.203	.897	.370
(E2) -10- (E6)	.293	.232	1.261	.207
(E3) -11- (E7)	.208	.221	.938	348
(E4) -12- (E8)	.270	.209	1.288	.198
(P) -21- (C)	.248	.169	1.467	.142

最后，我们还可以把正态化的残余值和期望的正态值来作图，如果观察的残余值和期望的残余值完全一致的，就是一条直线。从图 11.52 中可见，都在这条直线的附近。

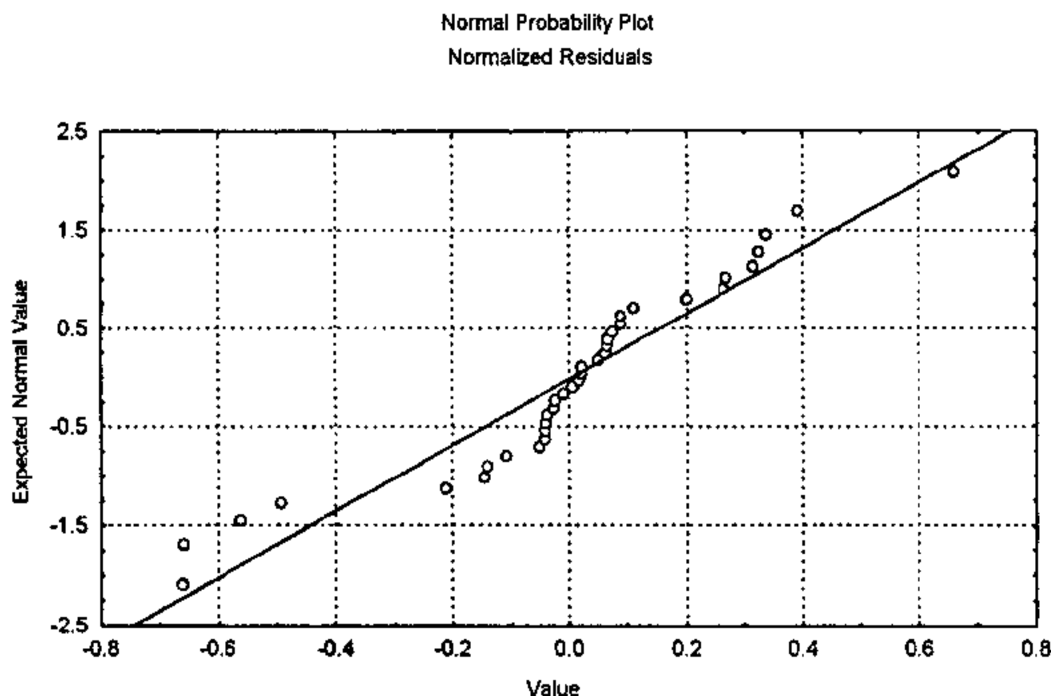


图 11.52

11.8. 抽样数目的估算

在上面的不同场合（如 6.1.7 和 8.2）里，我们都提到抽样，抽样是为了对总体作出推断，抽样数目的多少是一个关键的问题，但是究竟怎样决定抽样的数目，却还没有谈到，这里需要再作一些说明。

11.8.1. 决定样本数的几个因素

样本数的大小决定于几个因素；

1. 被调查事物总体的差异程度，这也就是标准差。如果所调查的事物是同质性的，需要的样本的数量可少一些；越是异质性，需要的样本越大。光是调查大学二年级的英语水平和调查全国各类学生的英语水平所需的样本数当然是不一样的。

2. 容许误差（D）的大小。容许误差越小，即抽样的平均误差（m）越小，把握程度（t）越大，抽样的数目就越大。反之，容许误差越大，即抽样的平均误差越大，把握程度就越小，而抽样的数目就越少。容许误差的大小取决于调查的目的、要求、经费和力量。例如我们用一个考试对一个样本的英语水平做调查，这个考试的满分是 100，抽样的平均误差定为 5 分（即±5 分内没有区别）和定为 10 分所需的样本数当然不一样。按照正态分布的原理，把

握程度和误差范围有关，如果我们把误差范围定为±1.96，在这个范围内的概率为95%，定为±2.58，则概率为99%。这就是我们在上面经常谈到的显著意义为0.05和0.01。为了计算的方便，我们往往把t定为2(95.45%)或3(99.73%)。容许误差(D=tm，例如t定为1.96，而m定为5分，则D=1.96×5 =9.8。

3. 抽取调查单位的抽样方法。基本的方法是简单的随机抽样，有重复抽样和不重复抽样之分。重复抽样指的是把抽出的样本放回取再抽，例如把一个六面的骰子掷两次，第一次是3，第二次再掷时，把3放回去再抽，因为骰子始终是六面。但在一个班里抽六个学生，抽了以后就没有必要再放回去再抽，这是不重复抽样。重复抽样所需的样本数比不重复抽样所需的样本数要多。一般的社会调查都是采取不重复抽样的方法，所以在下面的讨论中，我们只介绍不重复抽样的抽样数目的估算方法。除了简单的随机抽样方法外还有等距抽样、分层抽样和类型抽样等。每一种方法都有其计算样本数的公式。

11.8.2. 简单的随机抽样

简单的随机抽样的公式是：

$$n=\frac{t^2\sigma^2N}{\Delta^2N+t^2\sigma^2} \tag{11.46}$$

假定一个200人的年级，已知标准差为20，其方差便是400，t为1.96，如抽样平均误差为5，则D=1.96×5=9.8代入公式(11.46)，可得样本数为14.815。四舍五入即15人。换句话说，用15人的样本来推断200人的总体的把握程度的概率为95%。假定我们把概率提高到99%，t=2.58，而容许误差仍为9.8(这意味着抽样平均误差缩小到9.8/2.58=3.8)，样本数就要增加到24.348，即24人。

有时，我们有总体数，也作了抽样，也可以用下列公式来计算出样本的平均误差：

$$\mu=\sqrt{\frac{\sigma^2}{n}\ (1-\frac{n}{N})} \tag{11.47}$$

n是样本数，N是总体数，用这个公式计算出来的样本平均误差为5。这个计算样本的平均误差的公式是基本的公式，其他的抽样方法的平均误差的计算方法，都以此为基础，无非在计算方差时有所不同。

11.8.3. 分层抽样

分层抽样(stratified sampling)的抽样数并不决定于总体数，而是决定分为多少个层次，每个层次有多少数目。分层抽样有两种：一种是同等分配(equal distribution)，即每个层次的抽样数是一样的；另一种是按比例分配(proportional distribution)，即按每个层次的数目按比例决定抽样数。

同等分配样本数的公式是：

$$n=\frac{t^2 L\Sigma_{n=1}^{L} N_\eta^2 \sigma_\eta^2}{N^2\Delta^2+t^2\Sigma_{n=1}^{L} N_\eta \sigma_\eta^2} \tag{11.48}$$

按比例分配样本数的公式是：

$$n=\frac{t^2 L\Sigma_{n=1}^{L} N_\eta \sigma_\eta^2}{N^2\Delta^2+t^2\Sigma_{n=1}^{L} N_\eta \sigma_\eta^2} \tag{11.49}$$

如果是同等分配，L 的层次数，n 是抽样的总数，然后再 n/L，就可得出每一个层次的抽样数。

表 11.64

分层抽样的样本数

层次	N_h	σ_h^2	同等	按比例
1	100	75	16.82	7.72
2	150	110	16.82	11.58
3	200	150	16.82	15.44
4	225	200	16.82	17.37
5	250	275	16.82	19.31
总数	925		84.1	71.42

例如我们的总体分为 5 个层次，现决定 $t=1.96$，$\Delta=3$。用 (11.48) 公式，计算出应抽 84 个样本数，现有 5 个层次，因此每个层次应抽 84/5=16.8，四舍五入，即 17 个人。

如果是按比例分配，则应取得每个层次和总数的百分比，再乘以 n，例如第一层次为 100/925×71.42=7.72，其他层次，以此类推。

11.8.4. 整群抽样和系统抽样

整群抽样 (cluster sampling) 指的是按单位来抽样。有些时候，不大可能按个人来抽样，就只好先把总体分为不同类别 (m)，然后在每一种类别里抽若干单位。例如我们想了解北京市中学生的语文水平，设计了一个考试。如果我们用简单的随机抽样法来在全体中学生中抽样，将来的考试就比较难以组织。替换的办法是按地区来抽一些学校，让整个学校的学生都参加考试。分类抽样的公式如下：

$$m=\frac{Mt^2\sigma_b^2}{M\Delta_d^2+t^2\sigma_b^2} \tag{11.50}$$

m=每类的抽样数，M=类别数，σ_b^2=所有类别的方差，Δ_d^2=N/M（总体数/类别数，即每个类别总数）×Δ^2。假定 M 为 30，t 为 1.96，σ_b^2 为 100，Δ_d为 5，按 (11.50) 可算出 m 为 10.16，即每类抽 10 个单位即可。

系统抽样(systematic sampling)是根据总体数决定一个比例，再按照一定的排列(如成绩的高低或年龄的大小)等距进行抽样。系统抽样也叫等距抽样，例如我们决定抽 1/10 的学生来了解其语文水平，在 100 人的一个年级里就抽 10 人，可按其原来的水平从高到低排队，然后随机定一个数字，如 3，就每隔 10 个抽样，即 3、13、23、33 等。这个种办法已经决定抽样数目，无需再推算。但要计算其抽样的平均误差，一般就采用简单随机抽样计算误差的公式即可 (11.47)，如方差为 100，则样本平均误差为 $\sqrt{100/10\times(1-10/100)}=3$。

11.9. 小结

我们一开始就说过，统计学是一门专门的学问，不可能用短短的篇幅来覆盖，但在讨论语言学方法论时，又少不了它。故我们只涉及一些基本的概念和用法，既不谈它的推导原理，也不谈具体的运算。我们花了一些篇幅来介绍计算机程序运算，是因为它易学上手。由于计算机能够作从简单到复杂的各种统计，只要掌握基本操作方法，都能运算，这也容易产生另一个问题：在没有真正吃透统计学的精神，就到处使用，从滥用到误用。滥用和误用比不用还要危险，因为它罗列的一些使用不当的数据，更容易产生误导的作用。Reichmann (1961) 在他的《统计学的使用和滥用》里认为，统计学的时代已经来临，人类各种活动和自然界的各种现象都在使用统计学方法来测量并且作出解释。但是这些解释有时是明智的，有时却是愚蠢的。因为对统计学的看法存在着两种对立而又极端的观点：一种是现成的统计学有具体规定的含义，一贯正确，应该无条件地接受；另一种是现在什么东西都用统计学去证明，而事实上是什么都证明不了。这两种极端的观点都是同样荒谬的。引起这些想法的根子是报章杂志和广告里滥用了很多关于显著意义的虚假的陈述，既不说明数据是怎样来的，也不说明应该怎样正确理解显著意义。

所以我们应该看到统计学既非万能的，也不可能是一贯正确的，但是正确地使用统计学使我们能够解释数据，取得有意义的结果，而关键的问题在于懂得在什么时候使用什么正确的方法。要做到这一点，必须明确：

1. 你为什么要使用统计方法？你用它来描写你所收集的数据呢？还是用它来验证你的假设、预测、模型？还是用它来探索和了解事物的型式和内部关系？换句话说，你要用统计方法来达到什么目的？统计方法在你的研究中占什么地位？

2. 你有些什么数据？是范畴性的数据？还是连续性的数据？这些数据能否转换？这些数据适宜于使用什么统计方法？

3. 你能采用什么统计方法？你是否具备和熟悉你准备使用的统计软件？

4. 你期望会得到什么结果，这些结果能否满足你的要求和目的？

下面将从使用统计方法的不同目的和所需要的什么数据来加以归纳：

1. 归纳、描写和列表

你要做些什么？	怎样做？	你能得到什么？
1. 对连续性变量作归纳和测量/对其分布状况作图	使用简单描写统计学（交叉列表、频率分布等）	平均值、标准差、方差、偏态值、峰值；直方图、圆形图、散布图等
2. 非参数统计和分布模型的拟合	正态分布、二项分布分析、泊松分析等	众数、中位数、四分位数、四分位距、百分位、全距、绝对平均差等

续前表

你要做些什么？	怎样做？	你能得到什么？
3. 按样本、组、变量进行描写统计	所有上述有关的归纳统计	所有上述有关的归纳统计，但按样本、组、变量来整理
4. 对范畴性数据（如性别、职业等）进行列表、计算频率和百分比	频率分布，对两个和更多的变量交叉列表以描写其联合分布	χ^2 值，显著意义

2. 检验数据是否符合假设和预测

你要做些什么？	怎样做？	你能得到什么？
1. 某一种分布模型的估算或拟合	正态分布、χ^2 分布、t 分布、F 分布、二项分布、泊松分布等	显著意义、预测频率和实际频率的差别，残差
2. 两个或两个以上的组和样本的差别	(1) 比较两组的平均值和方差，t 检验和 f 检验	t 值，f 值，显著意义
	(2) 比较两个以上的组和样本的平均值和方差，ANOVA，MANOVA，判别分析	f 检验，显著意义，交互作用效应，判别函数，判别分
	(3) 两个或两个以上的组的分布的非参数比较，Kolmogorov-Smirnov 两样本检验	显著意义
3. 变量之间的差别	比较两个或更多的变量的平均值，t 检验，ANOVA	t 值，f 值，交互作用效应，显著意义
4. 变量之间的关系	(1) 两个连续性变量的简单线性关系：双向相关，线性回归	相关系数，显著意义，回归方程，预测值
	(2) 两个范畴性变量的简单线性关系：双向列联表	χ^2 值，显著意义
	(3) 连续性变量之间的多重线性关系：多元回归分析	f 值，相关系数，回归方程，预测值，残余值
	(4) 范畴性变量之间的多重线性关系：对数线性模型	交互作用效应
	(5) 非线性关系：非线性回归分析	回归方程

续前表

你要做些什么？	怎样做？	你能得到什么？
5. 不同组的变量之间的差别	(1) 协方差分析 (ANCOVA) (2) 分层线性或非线性回归	显著意义 回归方程
6. 变量（显露的和潜在的）之间的关系的模型	验证性因子分析，路径分析，结构方程模型	显著意义，参数值估算值，t 值，χ^2 值

3. 探索数据和寻找结构/型式/因子/聚类

你要做些什么？	怎样做？	你能得到什么？
1. 多重连续性变量的因子和维度	因子分析，多维度量表	主要成分分析，旋转方法，特征值
2. 变量或个案的聚类和“自然”分组	聚类分析	树形图
3. 跨时间观察的型式和倾向	时间系列	回归方程
4. 多重交叉量表的关系	对数线性模型	交互作用

读者不难发现，我们并没有完全介绍表中所列的各种统计方法，这主要是考虑到有些方法的通用性不太广，而且本书的篇幅有限。同时也应该指出的是我们使用统计方法的目的往往不止一个，例如我们既要整理和描写数据，但也想利用数据来验证我们的假设；我们既要整理和描写数据，还想了解这些数据之间有些什么内在的关系或型式等等，因此有些分析既是描写性的，也是验证性的或探索性的。

参考书目

Alderson, J. 1985. *Cognitive Psychology and Its Implications*. NY: W. H. Freeman

Allen, J. & Davies, A. 1977. *Testing and Experimental Methods*. Oxford: OUP

Allen, J. 1979. *A Plan-based Approach to Speech Act Recognition*. Technical report 131, University of Toronto, Toronto

Allen, J. P. B. & Davies, A. 1977. *Testing and Experimental Methods*. London: OUP

Anastasi, A. 1982. *Psychological Testing*. 5th Edition. New York: Macmillan

Bachman, L. 1990. *Fundamental Considerations in Language Testing*. Oxford: OUP

Bloomfield, L. 1927. Literate and illiterate speech. *American Speech*. 2: 423-439 (见 Hymes, D (ed.). 1964. *Language in Culture and Society*. pp. 391-396. New York: Harper and Row)

Bloomfield, L. 1925. Why a linguistic society? *Language*. 1: 1-5

Bloomfield. L. 1933. *Language*. Chicago: University of Chicago Press.

Boas, F. 1911. Introduction to *The Handbook of American Indian Languages*. Bureau of American Ethnology, Bulletin 40

Bogdan, R. C. & Biklen, S. K. 1982. *Qualitative Research in Education*. Boston: Allyn & Bacon

Bourhis, R. & Giles, H. 1976. The language of cooperation in Wales: a field study. *Language Sciences*. 42: 13-16

Bowers, G. & Clapper, J. 1989. Experimental methods in cognitive science. In Posner (ed.) 1989

Bowerman, M. 1974. Learning the structure of causative verbs: A Study in the relationship of cognitive, semantic and syntactic development. In E. Clark (ed.). *Papers and Reports on Child Language Acquisition* No. 8. Stanford University Committee on Linguistics, 142-178

Bresnan, Joan. 1978. A realistic transformational grammar. In M. Halle, J. Bresnan and J. Miller (eds.), *Linguistic Theory and Psychological Reality*. Cambridge, Mass: The MIT Press, 1-59

Broadbent, D. 1954. The role of auditory localisation in attention and memory span. *Journal of Experimental Psychology*. 47

Brown, J. 1988. *Understanding Research in Second Language Acquisition*. Cambridge: Cambridge University Press

Brown, R. and C. Hanlon. 1970. Derivational complexity and order of acquisition in child speech. In J. R. Hayes (ed.). *Cognition and the Development of Language*. New York: Wiley

Bulter, C. 1985. *Statistics in Linguistics*. Oxford: Basil Blackwell

Burroughs, G. E. R. 1971. *Design and Analysis in Educational Research*. Birmingham: University of Birmingham

Butts, R. E. and J. Hintikka. 1977. (eds.). *Basic Problems in Methodology and Linguistics: Logic, Methodology and Philosophy of Science*. Dordrecht: Reidel

Campbell, D. & Fiske, D. 1959. Convergent and discriminant validation of multitrait-multimethod matrix. *Psychological Bulletin*. 56: 81-105

Campbell, D. & Stanley, J. 1963. *Experimental and Quasi-experimental Designs for Research*. Reprinted from *Handbook of Research on Teaching*. Boston: Houghton Mifflin Company

Carberry, M. 1988. Pragmatic modeling: Toward a robust natural language interface. *Computational Linguistics*. 3

Carmines, E. & Zeller, R. 1979. *Reliability and Validity Assessment*. London: SAGE Publications

Carol, J. B. and M. N. White. 1973. Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*. 12: 563-576

Carroll, J. et al. 1971. *Word Frequency Book*. Boston: Houghton Mifflin Co. & NY: American Heritage Publishing Co., Inc.

Chafe, W. (ed.) 1980. *The Pear Stories: Cognitive, Cultural and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex

Charniak, E. 1977. A framed painting: The representation of commonsense knowledge fragment. *Cognitive Science*. 1 (4)

Child, D. 1970. *The Essentials of Factor Analysis*. London: Holt, Rinehart & William

Chomsky, N. 1957. *Syntactic Structure*. The Hague: Mouton

Chomsky, N. 1959. A review of B. F. Skinner's *Verbal Behavior*. *Language*. 35 (9): 26-58

Chomsky, N. 1964. *Current Issues in Linguistic Theory*. The Hague: Mouton

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press

Chomsky, N. 1969. Some empirical assumptions in modern philosophy of language. In S. Morgenbesser, P. Suppes and M. White (eds.). *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. New York: St. Martin's Press

Chomsky, N. 1969. Knowledge of Language. In K. Gunderson, and G. Maxwell (ed.). 1975. *Language, Mind and Knowledge*. Minneapolis: University of Minnesota Press

Chomsky, N. 1975. *Reflections on Language*. New York: Pantheon

Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris

Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass: The MIT Press

Chomsky, N. 1986a. *Knowledge of Language: Its Use, Origins, and Use*. New York:

Praeger

Chomsky, N. 1986b. *Barriers*. Cambridge, Mass: The MIT Press

Chomsky, N. 1987. Generative grammar: Its basis, development and prospects. *Studies in English Linguistics and Literature Special Issue*. Kyoto University of Foreign Studies

Chomsky, N. 1988. Linguistics and adjacent fields: the state of the art— a personal view. Lecture given in Israel

Chomsky, N. 1989. Some notes on economy of derivation and representation. ms. MIT

Chomsky, N. 1995. *A Minimalist Program for Linguistic Theory*. Gambridge, Mass: MIT Press

Chomsky, N. and Mitsou Ronat. 1979. *Language and Responsibility*. New York: Pantheon

Clark, H. 1973. Space, time, semantics, and the child. In E. Moore (eds.). *Cognitive Development and the Acquisition of Language*. New York: Academic

Clark, H. 1970. The primitives nature of children's relational concepts. In J. R. Hayes, (ed.). *Cognition and the Development of Language*. New York: Wiley

Cohen, R. 1987. Analyzing the structure of argumentative discourse. *Computational Linguistics*. 13 (1-2)

Cook, V. 1986. *Experimental Approaches to Second Language Learning*. Oxford: Pergamon

Cooper, R. & Fishman, J. 1974. The study of language attitudes. *International Journal of the Sociology of Language*. 3: 5-19

Cooper, R. & Weekes, A. 1983. *Data, Models and Statistical Analysis*. Oxford: Philip Alan

Crystal, D. 1987, *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press

Crystal, D. & Davy, D. 1969. *Investigating English Styles*. London: Longman Group

Curtiss, S. 1981. Dissociation between language and cognition. *Journal of Autism and Developmental Disorders*. 11: 55-130

Deets, J. 1967. *Invitation to Archaeology*. Garden City, NY: Natural Hisotry Press

Dennis, M. and Harry A. Whitaker. 1976. Language acquisition following hemidecortication: linguistic superiority of the left over right hemisphere. *Brain and Language*. 3: 404-33.

Donders, F. 1969. On the speed of mental processes (1868-1869). *Acta Psychologica*. 30: 412-431

Dowty, David. 1982. Grammatical relations and Montague grammar. In P. Jacobson and G. K. Pullum (eds.). *The Nature of Syntactic Representation*. Dordrecht: Reiddel. 79-130

Dressler, W. U. 1978. *Current Trends in Textlinguistics*. Berlin: Walter de Gruyter

Dundes, A. 1962. From etic to emic units in the structural study of folktales. *Jounral of American Folklore*. 75: 95-105

Eckardt, B. 1993. *What is Cognitive Science*. Cambridge, Mass: The MIT Press

El-Dash, L. & Tucker, R. 1975. Subjective reactions to various speech styles in Egypt. *International Journal of the Sociology of Language*. 6: 33-54

Eyesenck, M. 1988. *A Handbook of Cognitive Psychology.* London: Lawrence Erlbaum Associates

Farringdon, M. & J. 1987. A computer-aided study of the prose style of Henry Fielding and its support for his translation of the Military History of Charles XII. In Ager, D. et al (eds.). *Advances in Computer-aided Literary and Linguistic Research.* Birmingham: University of Aston in Birmingham

Fasold, R. 1984. *The Sociolinguistics of Society.* Oxford: Basil Blackwell

Ferguson, C. 1966. National sociolinguistic profile formulas. In Bright, W. (ed.) *Sociolinguistics.* The Hague: Mouton. pp. 309-324

Fikes, R. & Nilsson, N. 1971. STRIPS: a new approach to the application of theorem proving a problem solving. *Artificial Intelligence.* 2 : 189-208

Fillmore, C. L. 1967. The case for case. In E. Bach and R. T. Harns (eds.). *Universals in Linguistic Theory.* New York: Holt, Rinehart & Winston

Firbas, J. 1975. On the thematic and non-thematic section of the sentences. Ringbom (ed.) *Style and Text: Studies Presented to Nils-Erik Enkvist.* Stockholm: Skriptor

Firth, J. 1957. Ethnographic analysis and language with reference to Malinowski views. In J. R. Firth (ed.). *Man and Culture. An Evaluation of the Work of Bronislw Malinowski.* London: Routledge & Kegan Paul. pp. 93-118

Fishman, J. 1968. Sociolinguistics and the language problems of developing countries. In Fishman, Ferguson, and Das Gupta (eds.). *Language Problems of Developing countries.* New York: Hohn Wiley and Sons

Flint, H. 1979. Stable societal diglossia in Norfolk Island. In Mackey & Ornstein (eds.). *Sociolinguistic Studies in Language Contact: Methods and Cases.* The Hague: Mouton

Foss, B. (ed.) 1966. *Introduction in New Horizons in Psychology.* Harmondsworth: Penguin.

Frake, C. 1968. The ethnographic study of cognitive systems. In Fishman, J. (ed.). *Readings in the Sociology of Language.* The Hague: Mouton. pp. 434-446

Fries, C. 1940. *American English Grammar.* New York: Appleton-Century-Crofts

Fries, C. 1952. *The Structure of English.* New York: Harcourt, Brace and Company

Gagne, R. 1977. *The Conditions of Learning.* NY: Holt, Rinehart & Winston

Garside, R. et al (eds.). 1987. *The Computational Analysis of English.* London: Longman

Garside, R. & Leech, Fanny. 1987. The UCREL probabilistic parsing system. In Garside et al (eds) 1987.

Gazdar, Gerald. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry.* 12 : 155-184

Gazdar, Gerald, Edwan Klein, Gerffrey K. Pullum, and Ivan Sag 1985. *Generalized Phrase Structure Grammar.* Cambridge, Mass: Harvard University Press

Good, I. 1953. On the population frequencies of species and the estimation of population parameter. *Biometrika.* 40 : 237-264

Green, P. 1975. *The Language Laboratory in School*. Edinburgh: Oliver & Boyd

Greenbaum, S. & Quirk, R. 1970. *Elicitation Experiments in English Linguistics Studies in Use and Attitude*. London: Longman Group

Greenberg, J. 1956. The measurement of linguistic diversity. *Language*. 32 (2): 109-115

Greenberg, J. 1963. Some universals of gdrammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.). *Universal of Language*. Cambridge, Mass: MIT Press. 58-90

Greenberg, J. 1968. *Anthropological Linguistics*. New York: Random House

Greenberg, J. 1978. Typology and cross-linguistic generlizations. In J. B. Greenberg (ed.). *Universals of Human Language: Volume 1: Method and Theory*. Stanford: Stanford University Press

Grosz, B. 1977. The representation and use of focus in a system of understanding dialog. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*. Cambridge, MA

Grosz, B. et al. 1989. Discourse. In Posner (ed.) 1989

Gumperz, J. J. 1967. The social setting of linguistic behaviour. 见 Dittmar, N. 1976. *Sociolinguistics: A Critical Survey of Theory and Application*. London: Edward Arnold. pp. 190-191

Gumperz, J. 1977. Sociocultural knowledge in conversational inference. In Muriel Saville-Troike (ed.). *Linguistics and Anthropology*. Washington, D. C.: Georgetown Unversity Press. pp. 191-192

Gumperz, J. J. & Hymes, D. (eds.). 1986. *Directions of Sociolinguistics: Ethnography of Communication*. Oxford: Basil Blackwell

Halliday, M. A. K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold

Halliday, M. A. K. 1985. *An Introduction to Functional Linguistics*. London: Longman

Harris, R. (ed.). 1988. *Linguistic Thought in England, 1914~1945*. NY: Routledge

Harris, Z. 1951. *Methods of Structural Linguistics*. Chicago: University of Chicago Press

Harris, Z. 1960. Preface to the Phoenix edition of *Structural Linguistics*. Hartmann, P. (1963). Theorie der Sprachwissenschaft. Assen: van Gorcum

Hatch, E. & Farhady, H. 1982. *Research Design and Statistics for Applied Linguistics*. NY: Newbury House

Haugen, E. (1950) 其中文译文《现代语言学的方向》见岑麒祥 (1992):《国外语言学论文选译》，北京：语文出版社。

Hawkins, C. & Weber, J. 1980. *Statistical Analysis*. New York: Harper & Row

Herden, G. 1960. *Type-Token Statistics*. The Hague: Mouton

Herden, G. 1964. *Quantitative Linguistics*. London: Butterworth

Hjemslev, L. 1969. *Prolegomena to a Theory of Language*, Madison: University of Wisconsin Press

Ho Sai Keung & Hoosain, R. 1984. Hemisphere differences in the perception of Chinese op-

posites. In Kao, H. & Hoosain, R. (eds.) 1984

Hobbs, J. 1979. Coherence and co-reference. Cognitive Science. 1 : 67-82

Hope, J. 1994. *The Authorship of Shakespeare's Plays: A Socio-linguistic Study*. Cambridge: Cambridge University Press

Horvath, B. 1985. *Variation in Australian English*. Cambridge: CUP

Huang, C. -T. James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. Doctoral dissertation, MIT

Hymes, D. 1964. Introduction to the scope of linguistic anthropology. In Hymes (ed.) *Language in Culture and Society*. NY: Harper & Row

Hymes, D. 1972. On communicative competence. In Pride and Holmes (eds). *Sociolinguistics*. Harmondsworth: Penguin

Hymes, D. 1972, Models of the interaction of language and social life. In Gumperz, J. & Hymes, D. (eds). *Directions in Sociolinguistics: the ethnography of communication*. New York: Holt, Rinehart & Winston. 56-57

Hymes, D. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*. Philadelphia: University of Pennsylvania Press

Jöreskog, K. & Söbom, D. 1989. *Lisrel 7: A Guide to the Program and Applications*. 2nd Edition. Chicago: Jöreskog & Söbom/SPSS Inc.

Jöreskog, K. & Söbom, D. 1989. *Lisrel 7: User's Reference Guide*. Scientific Software, Inc.

Jackendoff, R. 1972. *Interpretation in Generative Grammar*. Cambridge, Mass: The MIT Press

Jackendoff, R. 1977. *X'-syntax: A Study of Phrase Structure*. Cambridge, Mass: The MIT Press

Jacob, Evelyn 1987. Qualitative research traditions: A review. *Review of Educational Research*. 57 (1): -50

Jacobson, R., Fant, C., & Halle, M. 1951. *Preliminaries to Speech Analysis—The Distinctive Features and Their Correlates*. 中译文载《国外语言学》1981, 3-4。

Johnson-Laird, P. 1988. *The Computer and the Mind*. London: Fontana

Just, M. & Carpenter, P. 1980. A theory of reading: from eye fixations to comprehension. *Psychological Review*. 87 : 329-352

Just, M. & Carpenter, P. 1987. *The Psychology of Reading and Comprehension*. Newton, MA: Allyn and Bacon

Kamenyama, Megumi. 1984. Subjective/logophoric bound anaphor Zibun. In J. Drogo, V. Mishra, and D. Testen (eds.). *Papers from the 20th Regional Meeting*. Chicago Linguistics Society, Chicago, 228-238

Kao, H. & Hoosain, R. (eds.) 1984. *Psychological Studies of of Chinese Language*. Hong Kong: Chinese Language Society of Hong Kong

Kaplan, Ronald and Joan Bresnan 1982. Lexical-functional grammar: A formal system for grammatical representation. In Bresnan, J. (ed.). Mental Represention and Grammatical

Relations. Cambridge, Mass: MIT Press

Kasher, A. ed. 1991. *The Chomskyan Turn.* Cambridge, Mass: MIT Press

Kinsella, V. 1982. *Cambridge Language Teaching Survey.* Vol. 1 & Vol. 2. Cambridge: Cambridge University Press

Kinsella, V. 1985. *Cambridge Language Teaching Survey.* Vol. 3. Cambridge: Cambridge University Press

Kloss, H. 1968. Notes concerning a language-nation typology. In Fishman et al. 1968. *Language Problems of Developing Nations.* N. Y.: John Wiley & Sons

Kuhn, T. 1970. *The Structure of Scientific Revolution.* Chicago: University of Chicago Press

Kuo, E. 1979. Measuring communicativity in multilingual societies: the case of Singapore and West Malaysia. *Anthropological Linguistics.* 21 (7): 328-40

Labov, W. 1966. *The Social Stratifification of English.* Washington, D. C.: Center for Applied Linguistics

Labov, W. 1970. The study of language in its social context. *Stadium Generale.* 23. 30-87 See also Fishman, J. 1976. *Advances in the Sociology of Language.* Vol. 1. The Hague: Mouton

Labov, W. 1972. *Sociolinguistic Patterns.* Philadelphia: University of Pennsyvanlia Press.

Lambert, W. et al. 1960. Evaluative reactions to spoken language. *Jounal of Abnormal and Social Psychology.* 60: 44-51

Larsen -Freeman, D. & Long, M. 1991. *An Introduction to Second Language Acquisition Research.* London: Longman

Laughlin, R. 1980. Of shoes and ships and sealing wax: sudries from Zinacatán. *Cotributions to Anthropology* No. 25. Washington, DC: Smithsonian Institution

Leech, G & Short, M. 1981. *Style in Fiction.* London: Longman Group

Leech, G. 1977. *Semantics.* Harmondsworth: Penguin

Lenneberg, E. H. 1967. *Biological Foundations of Language.* New York: Wiley

Lieberson, S. 1964. An extension of Greenberg's linguistic diversity measures. *Language*, 40: 526-31

Lieberson, S. 1967. Language questions in census. In Lieberson (ed.). *Explorations in Sociolinguistics.* 134-51

Liberson, S. & Hansen, L. 1974. National developmemt, mother tongue diversity, and the comparative study of nations. *American Sociological Review.* 39: 523-541

Lord, F. 1980. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum

Lyons, J. 1968. *Introduction to General Linguistics.* Cambridge: Cambridge University Press

Malinowski, B. 1922 (1961). *Argonauts of the Western Pacific* New York: Dutton

Malinowski, B. 1935, *Coral Gardens and Their Magic.* London: G. Allen & Unwin

Mandelbrot, B. 1965. Information theory and psycholinguistics. In Oldfield, R. & Marshall,

J. (eds.) 1968. *Language*. Harmondsworth: Penguin

Marshall, Ian. 1987. Tag selection using probabilistic methods. In Garside et al (eds.) 1987

McConnell, G. 1979. Constructing language profiles by polity. In Mackey & Ornstein (eds.). *Sociolinguistic Studies in Language Contact: Methods and Cases*. The Hague: Mouton

McKeown, K. 1985. *Text Generation*. New York: Cambridge University Press

Meillet, A. 1924. La methode comparativ en linguistique historique. Paris. 中译本见岑麒祥译《国外语言学论文选译》，语文出版社，1992，p. 18

Mellish, C. 1982. *Incremental Evaluation: An approach to the Semantic Interpretation of Noun Phrases*. Technical Report, University of Sussex Cognitive Studies Programme, Sussex, Engl.

Miles, M. & Huberman, M. 1984. *Qualitative Data Analysis*. London: Sage

Milroy, L. 1987. *Observing & Analysing Natural Language*. Oxford: Blackwell

Mittins, et al. 1970. *Attitudes to English Usage*. London: Oxford University Press

Montague, Richard. 1974. *Formal Philosophy: Selected Papers of Richard Montague*. R. Thomason (ed.). New Haven: Yale University Press

Mosteller, F. & Wallace, D. 1985. Deciding authorship. In Tanur et al (eds.). *Statistics: A Guide to the Unknown*. Monterey, Cal.: Wadsworth & Brooks/Cole

Mosteller, F. & Wallace, D. 1964. *Inference and Disputed Authorship: The Federalist*. Reading, Mass.: Addison-Wesley

Newmeyer, Frederick. 1980. *Linguistic Theory in America: The First Quarter-Century of Transformational Generative Grammar*. New York: Academic Press

Newmeyer, Frederick. 1983. *Grammatical Theory: Its Limits and Its Possibilities*. Chicago: University of Chicago Press

Newmeyer, F. 1988. *Linguistics: The Cambridge Survey*. Volumes 1-4. Cambridge: Cambridge University Press

Norusis, M. 1985. *SPSS: Advanced Statistics Guide*. Chicago: SPSS Inc.

Osgood, C., Suci, C. & Tannenbaum, P. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press

Osherson, D., Kosslyn, S. & Hollerbach, J. 1990. *Visual Cognition. An Invitation to Cognitive Science*. Vol. 2. Cambridge, Mass: The MIT Press

Paivio, A. & Begg, J. 1981. *Psychology of language*. Englewood Cliffs, NJ: Prentice Hall.

Patton, Michael. 1990. *Quatitative Evaluation and Research Methods*. 2nd edition. Beverly Hills, CA: Sage

Pearson, B. 1977. *Introduction to Linguistic Concepts*. NY: Alfred Knopf, Inc.

Perlmutter, David and Carol Rosen. (eds.). 1984. *Studies in Relational Grammar* 2. Chicago: University of Chicago Press.

Petöfi, J. 1975. *Beyond the sentence, between linguistics and logic*. Ringbom (ed.) 1975

Pike, K. 1967. *Language in Relation to a Unified Theory of the Structure of Human Behav-*

inur. The Hague: Mouton

Polanyi, M. 1986. A theory of discourse structure and discourse coherence. In W. Eilfort, et al (eds.). *Proceedings of the 21st Reginal Meeting of the Chicago Linguistic Society*. Chicago: University of Chicago Press

Posner, M. (ed.) 1989. *Foundations of Cognitive Science*. Cambridge, Mass: The MIT Press

Quine, W. V. O. 1972. Methodological reflections on current linguistic theory. In Donald Davison and Gilbert Harman (eds.). *Semantics of Natural Language*. New York: Humanities Press

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. 1972. *A Grammar of Comtempory English*. London: Longman

Reichmann, W. 1961. *Use and Abuse of Statistics*. Harmondsworth, Middlesex: Penguin

Reid, E. 1978. Social and stylistic variation in the speech of children: some evidence from Edinburgh. In Trudgill (ed.). 1978. *Sociolinguistic patterns in British English*. London: Arnold. pp. 158-172

Richards, J. 1985. *Longman Dictionary of Applied Linguistics*. London: Longman

Rieser, H. 1978. On the development of text grammar. In Dressler, W. (ed.)

Robin, R. H. 1989. *General Linguistics*. 4th Edition. London: Longman

Romaine, S. 1978. Post-vocal /r/ in Scottish English. In Trudgill (ed.) 1978. *Sociolinguistic patterns in British English*. London: Arnold. pp. 144-57

Roshenshine, B. 1976. Recent research on teaching behaviors and student achievement. *Journal of Teacher Education*. 27: 61-64

Rummel, R. 1970. *Applied Factor Analysis*. Evanston: Northwestern University Press

Sag, Ivan. 1986. Grammatical hierarchy and linear precedence. *CSLI Report*, No 60, Stanford University

Sampson, G. 1987. The grammatical database and parsing scheme. In Garside et al (eds.)

Sankoff, G. 1980 . A quantitative paradigm for the study of communicative competence. In Sankoff (ed.). *The Social Life of Language*. Philadelphia: University of Pennsyvanlia Press. pp. 47-79

Santa, J. 1977. Spatial transformations of words and pictures. *Journal of Experimental Psychology: Human Learning and Memory*. 3: 418—427

Sapir, E. 1921. *Language*. NY: Harcourt, Brace & World

Saussure, F. 1966. *Course in General Linguistics* (transl. Wade Baskin), NY: McGraw-Hill

Saville-Troike, M. 1989. *The Ethnography of Communication*. 2nd Edition. Oxford: Basil Blackwell

Schmidt, S. 1978. Some problems of communicative text theories. In Dressler (ed.)

Seliger, H. & Long, M. 1983. *Classroom Oriented Research in Second Language Acquisition*. Cambridge, Mass.: Newbury House

Seliger, H. & Shohamy, E. 1989. *Second Language Research Methods*. London: OUP

Shuy, R. W., Wolfram, W. A. & Riley, W. K. 1968. *Field Techniques in an Urban Language Study.* Washington, D. C.: Center for Applied Linguistics.

Simpson, Jane 1983. *Aspects of Warlpiri morphology and Syntax.* Doctoral dissertation, MIT.

Slobin, D. 1979. *Psycholinguistics.* 2nd Edition. Glenview, Ill: Scott, Foresman and Co.

Smith, Jr., P. 1970. *A Comparison of the Cognitive and Audiolingual Approaches to Foreign Language Instruction: The Pennsylvania Foreign Language Project.* Pennsylvania: The Centre for Curriculum Development, Inc.

Smith, N., and I. M. Tsimpli. 1991. Linguistic modularity? A case study of a 'Savant' linguist. *Lingua.* 84: 315-351

Statistica for Windows. 1995 (*Computer Program Manual*). Tulsa, OK: StatSoft, Inc.

Stern, H. 1983. *Fundamental Concepts of Language Teaching.* Oxford: OUP

Sternberg, S. 1969. The discovery of processing stages: Extensions of Donders's method. *Acta Psychologica.* 30: 276-315

Stewart, W. 1962. An outline of linguistic typology for describing multilingualism. In Rice. F. 1962. *Study of the Role of Second Language in Asia, Africa and Latin America.* Washington, D. C.: Center for Applied Linguistics. pp. 15-25

Stewart, W. 1968. A sociolinguistic typology for describing national multilingualism. In Fishman. 1968. *Reading in the Sociology of Language.* The Hague: Mouton. pp. 531-545

Swinney, D. 1979. Lexical access during sentence comprehension: Reconsideration of context effects. *Journal of Verbal Learning and Verbal Behavior.* 18: 645-659

Szymura, J. 1988, Bronislaw Malinowski's 'Ethnographic Theory of Language'. In Nancy P. Hickerson (ed.). *Linguistic Anthropology.* NY: Holt, Rinehart & Winston

Tanenhaus, M. 1988. Psycholinguistics: an overview. In Newmeyer, F. (ed.) *Linguistics: The Cambridge Survey.* Vol. III.

Tannen, D. 1981. Indirectness in discourse: ethnicity as conversation style. *Discourse Process.* 4 (3): 221-38

Thorndike, R. 1977. *Measurement and Evaluation in Psychology and Education.* NY: Wiley and Sons

Townsend, J. 1953. *Introduction to Experimental Method.* New York: McGraw Hill.

Treisman, A. 1960. Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology.* 12

Trudgill, P. 1983. *On Dialect.* Oxford: Blackwell

Trudgill, P. & Tzavaras, G. 1977. Why Albanian-Greeks are not Albanians: language shift in Attica and Biotia. In Giles, H. (ed.). *Language, Ethnicity and Intergroup Relations.* London: Academic Press

Tsuda, Aoi. 1984. *Sale Talk in Japan and the United States.* Washington, DC: Georgetown University Press

Tuckman, B. 1978. *Conducting Educational Research*. 2nd Ed. New York: Harcourt Brace Jovanovich

Utan, R. 1969. Some general characteristics of interrogative systems. *WPLU*. 1. 41-63

van Dijk, T. & Kintsch, W. 1978. Cognitive psychology and discourse: recalling and summarizing stories. In Dressler (ed.) 1978

Webber, B. 1987. The interpretation of tense in discourse. *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Stanford, CA. pp. 147-154

Weinreich, U. 1957. Functional aspects of Indian bilingualism. *Word*. 13 (2) : 203-33

Whitaker, A. 1970. Linguistic competence: Evidence from Aphasia. *Glossa*. 4: 46-54

Widdowson, H. 1979. *Explorations in Applied Linguistics*. Oxford: OUP

William, C. 1946. Yule's 'Characteristic' and the 'Index of Diversity'. *Nature*. 157 : 482

Williams, F. 1974. The identification of linguistic attitudes. *International Journal of the Sociology of Language*. 3 : 21-32

Williams, F. 1974. The identification of linguistic attitudes. *International Jorunal of the Sociology of Language*. 3 : 21-32

Winograd, T. 1972. *Understanding Natural Language*. New York: Academic Press

Winograd, T. 1982. *Language as a Cognitive Process: Syntax*. London: Addison-Wesley

Wolfson, N. & Manes, J. 1980. The compliment as a social stgrategy. *International Journal of Human Communication*. 13 (3)

Wolfson, N. 1981. Compliments in cross-cultrual perspective. *TESOL Quarterly*. 15 (2)

Wonnacott, R. & Wonnacott, T. 1985. *Introductory Statistics*. 4th Ed. New York: John Wiley & Sons

Wood, W. et al. 1972. *The Lunar Sciences Natural Language Information System: Final Report*. BBN Report 2378. Cambridge, Mass: Bold, Beranek and Newman

Woods, A., Fletcher, P. & Hughes, A. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press

Yamada, J. 1990. *Laura: a case for modularity of language*. Cambridge, Mass: MIT Press

Yule, G. 1968. *The Statistical Study of Literary Vocabulary*. New York: Archon

Zift, G. 1935. *The Psycho-Biology of Language*. Boston: Houghton Mifflin Co.

王孝玲，1986.《教育统计学》，上海：华东师范大学出版社

王初明，1990.《应用心理语言学》，长沙：湖南教育出版社

王初明、亓鲁霞，1992.《外语听力策略个案研究》，载桂诗春（主编）1992

冯志伟，1991.《数学与语言》，湖南：湖南教育出版社

朱德熙，1962.《句法结构》，《中国语文》，8-9 月号

朱德熙，1980.《汉语句法中的歧义现象》，《中国语文》，第 2 期

孙建一，1989.《现代汉语字频测定和分析》，载陈原（主编）1989

李绍山，1992，《影响中国学生英语阅读理解的主要因素》，载桂诗春（主编）1992

吴伟平，1994.《法律语言学：会议、机构和刊物》，《国外语言学》，第 4 期

陈原，（主编）1989.《现代汉语定量分析》。上海：上海教育出版社
林连书，1955.《英语实验研究方法 》，广州：中山大学学报编辑部
胡壮麟，1994.《语篇的衔接和与连贯》，上海：上海外语教育出版社。
桂诗春，1985 .《心理语言学》，上海：上海外语教育出版社
桂诗春，1986.《标准化考试—理论、原则与方法》. 广州：广东高等教育出版社
桂诗春，1988.《应用语言学》，长沙：湖南教育出版社
桂诗春，1993.《应用语言学与认知科学》，《语言文字应用》，第 3 期
桂诗春，1995.《从这个地方很郊区谈起》，《语言文字应用》，1995 年第 3 期。
桂诗春，1991.《实验心理语言学纲要》，长沙：湖南教育出版社
桂诗春(主编)，1992.《中国学生英语学习心理》，长沙：湖南教育出版社
桂诗春，1988a.《中国学生英语词汇量调查》、《我国英语专业英语词汇量的调查和分析》，载《应用语言学与中国英语教学》，济南：山东教育出版社
桂诗春，1988b.《我国英语专业学生社会心理分析》，载《应用语言学与中国英语教学》，济南：山东教育出版社
桂诗春，1991.《实验心理语言学纲要》，长沙：湖南教育出版社
桂诗春，1992.《中国英语学生的心理词汇研究》，载桂诗春（主编）1992
桂诗春，1992.《“外语要从小学起”质疑》，《外语教学与研究》，第 4 期

附录A　正态曲线的面积和纵线表

标准分 (1) $z=x/\sigma$	面积			纵线 (5) y
	(2) 自 $\overline{X}$ 至 z	(3) 较大部分	(4) 较小部分	
0.00	0.0000	0.5000	0.5000	0.3989
0.01	0.0040	0.5040	0.4960	0.3989
0.02	0.0080	0.5080	0.4920	0.3989
0.03	0.0120	0.5120	0.4880	0.3988
0.04	0.0160	0.5160	0.4840	0.3986
0.05	0.0199	0.5199	0.4801	0.3984
0.06	0.0239	0.5239	0.4761	0.3982
0.07	0.0279	0.5279	0.4721	0.3980
0.08	0.0319	0.5319	0.4681	0.3977
0.09	0.0359	0.5359	0.4641	0.3973
0.10	0.0398	0.5398	0.4602	0.3970
0.11	0.0438	0.5438	0.4562	0.3965
0.12	0.0478	0.5478	0.4522	0.3961
0.13	0.0517	0.5517	0.4483	0.3956
0.14	0.0557	0.5557	0.4443	0.3951
0.15	0.0596	0.5596	0.4404	0.3945
0.16	0.0636	0.5636	0.4364	0.3939
0.17	0.0675	0.5675	0.4325	0.3962
0.18	0.0714	0.5714	0.4286	0.3925
0.19	0.0753	0.5753	0.4247	0.3918
0.20	0.0793	0.5793	0.4207	0.3910
0.21	0.0832	0.5832	0.4168	0.3902
0.22	0.0871	0.5871	0.4120	0.3894
0.23	0.0910	0.5910	0.4090	0.3885
0.24	0.0948	0.5948	0.4052	0.3876
0.25	0.0987	0.5987	0.4013	0.3867
0.26	0.1026	0.6026	0.3974	0.3857
0.27	0.1064	0.6064	0.3936	0.3847
0.28	0.1103	0.6103	0.3897	0.3836
0.29	0.1141	0.6141	0.3859	0.3825
0.30	0.1179	0.6179	0.3821	0.3814
0.31	0.1217	0.6217	0.3783	0.3802
0.32	0.1255	0.6255	0.3745	0.3790
0.33	0.1293	0.6293	0.3707	0.3778
0.34	0.1331	0.6331	0.3669	0.3765

附录 *A*(续)

标准分	面积			纵线
(1) $z=x/\sigma$	(2) 自 $\overline{X}$ 至 z	(3) 较大部分	(4) 较小部分	(5) y
0.35	0.1368	0.6368	0.3632	0.3752
0.36	0.1406	0.6406	0.3594	0.3739
0.37	0.1443	0.6443	0.3557	0.3725
0.38	0.1480	0.6480	0.3520	0.3712
0.39	0.1517	0.6517	0.3483	0.3697
0.40	0.1554	0.6554	0.3446	0.3683
0.41	0.1591	0.6591	0.3409	0.3668
0.42	0.1628	0.6628	0.3372	0.3653
0.43	0.1664	0.6664	0.3336	0.3637
0.44	0.1700	0.6700	0.3300	0.3621
0.45	0.1736	0.6736	0.3264	0.3605
0.46	0.1772	0.6772	0.3228	0.3589
0.47	0.1808	0.6808	0.3192	0.3572
0.48	0.1844	0.6844	0.3156	0.3555
0.49	0.1879	0.6879	0.3121	0.3538
0.50	0.1915	0.6915	0.3085	0.3521
0.51	0.1950	0.6965	0.3050	0.3503
0.52	0.1985	0.6985	0.3915	0.3485
0.53	0.2019	0.7019	0.2931	0.3467
0.54	0.2054	0.7054	0.2946	0.3448
0.55	0.2088	0.7088	0.2912	0.3429
0.56	0.2123	0.7123	0.2877	0.3410
0.57	0.2157	0.7157	0.2843	0.3319
0.58	0.2190	0.7190	0.2810	0.3372
0.59	0.2224	0.7224	0.2776	0.3352
0.60	0.2257	0.7257	0.2743	0.3332
0.61	0.2291	0.7291	0.2709	0.3312
0.62	0.2334	0.7324	0.2676	0.3292
0.63	0.2357	0.7357	0.2643	0.3271
0.64	0.2389	0.7389	0.2611	0.3251
0.65	0.2422	0.7422	0.2578	0.3230
0.66	0.2454	0.7454	0.2546	0.3209
0.67	0.2486	0.7486	0.2514	0.3187
0.68	0.2517	0.7517	0.2483	0.3166
0.69	0.2549	0.7549	0.2451	0.3144

附录 *A*(续)

标准分 (1) $z=x/\sigma$	面积 (2) 自 $\bar{X}$ 至 z	面积 (3) 较大部分	面积 (4) 较小部分	纵线 (5) y
0.70	0.2580	0.7589	0.2420	0.3123
0.71	0.2611	0.7611	0.2389	0.3101
0.72	0.2642	0.7642	0.2358	0.3079
0.73	0.2673	0.7673	0.2327	0.3056
0.74	0.2704	0.7704	0.2296	0.3034
0.75	0.2734	0.7734	0.2266	0.3011
0.76	0.2764	0.7764	0.2236	0.2989
0.77	0.2794	0.7794	0.2206	0.2988
0.78	0.2823	0.7823	0.2177	0.2943
0.79	0.2852	0.7852	0.2148	0.2920
0.80	0.2881	0.7881	0.2119	0.2897
0.81	0.2910	0.7910	0.2690	0.2874
0.82	0.2939	0.7939	0.2061	0.2850
0.83	0.2967	0.7967	0.2033	0.2827
0.84	0.2995	0.7995	0.2005	0.2803
0.85	0.3023	0.8023	0.1977	0.2780
0.86	0.3051	0.8051	0.1940	0.2756
0.87	0.3078	0.8078	0.1922	0.2732
0.88	0.3106	0.8106	0.1894	0.2709
0.89	0.3133	0.8133	0.1867	0.2685
0.90	0.3159	0.8159	0.1841	0.2661
0.91	0.3186	0.8180	0.1814	0.2637
0.92	0.3212	0.8212	0.1788	0.2613
0.93	0.3238	0.8238	0.1762	0.2589
0.94	0.3264	0.8264	0.1736	0.2565
0.95	0.3289	0.8289	0.1711	0.2541
0.96	0.3315	0.8315	0.1685	0.2516
0.97	0.3340	0.8340	0.1660	0.2492
0.98	0.3365	0.8365	0.1635	0.2468
0.99	0.3389	0.8389	0.1611	0.2444
1.00	0.3413	0.8413	0.1587	0.2420
1.01	0.3438	0.8438	0.1562	0.2396
1.02	0.3461	0.8461	0.1539	0.2371
1.03	0.3485	0.8485	0.1515	0.2347
1.04	0.3508	0.8508	0.1492	0.2323

附录 A(续)

标准分	面积			纵线
(1) $z=x/\sigma$	(2) 自 $\overline{X}$ 至 z	(3) 较大部分	(4) 较小部分	(5) y
1.05	0.3531	0.8531	0.1469	0.2299
1.06	0.3554	0.8554	0.1446	0.2275
1.07	0.3577	0.8577	0.1423	0.2251
1.08	0.3599	0.8599	0.1401	0.2227
1.09	0.3621	0.8621	0.1379	0.2203
1.10	0.3643	0.8643	0.1357	0.2179
1.11	0.3665	0.8665	0.1335	0.2155
1.12	0.3686	0.8686	0.1314	0.2131
1.13	0.3708	0.8708	0.1292	0.2107
1.14	0.3729	0.8729	0.1271	0.2083
1.15	0.3749	0.8740	0.1251	0.2059
1.16	0.3770	0.8770	0.1230	0.2036
1.17	0.3790	0.8790	0.1210	0.2012
1.18	0.3810	0.8810	0.1190	0.1989
1.19	0.3830	0.8830	0.1170	0.1965
1.20	0.3849	0.8849	0.1151	0.1942
1.21	0.3869	0.8869	0.1131	0.1919
1.22	0.3888	0.8888	0.1112	0.1895
1.23	0.3907	0.8907	0.1093	0.1872
1.24	0.3925	0.8925	0.1075	0.1849
1.25	0.3944	0.8944	0.1056	0.1826
1.26	0.3962	0.6962	0.1038	0.1804
1.27	0.3980	0.8980	0.1020	0.1781
1.28	0.3997	0.8997	0.1003	0.1758
1.29	0.4015	0.9015	0.0985	0.1736
1.30	0.4032	0.9032	0.0968	0.1714
1.31	0.4049	0.9049	0.0951	0.1691
1.32	0.4066	0.9066	0.0934	0.1669
1.33	0.4082	0.9082	0.0918	0.1647
1.34	0.4099	0.9999	0.0901	0.1626
1.35	0.4115	0.9115	0.0885	0.1604
1.36	0.4131	0.9131	0.0869	0.1582
1.37	0.4147	0.9147	0.0853	0.1561
1.38	0.4162	0.9162	0.0838	0.1539
1.39	0.4177	0.9177	0.0823	0.1518

附录 *A*(续)

标准分	面积			纵线
(1) $z=x/\sigma$	(2) 自 $\bar{X}$ 至 z	(3) 较大部分	(4) 较小部分	(5) y
1.40	0.4192	0.9192	0.0808	0.1497
1.41	0.4207	0.9207	0.6793	0.1476
1.42	0.4222	0.9222	0.0778	0.1456
1.43	0.4236	0.9236	0.0764	0.1435
1.44	0.4251	0.9251	0.0749	0.1415
1.45	0.4265	0.9265	0.0735	0.1394
1.46	0.4279	0.9279	0.0721	0.1374
1.47	0.4292	0.9292	0.0708	0.1354
1.48	0.4306	0.9306	0.0694	0.1334
1.49	0.4319	0.9319	0.0681	0.1315
1.50	0.4332	0.9332	0.0868	0.1295
1.51	0.4345	0.9345	0.0655	0.1276
1.52	0.4357	0.9357	0.0643	0.1257
1.53	0.4370	0.9370	0.0660	0.1238
1.54	0.4382	0.9382	0.0618	0.1219
1.55	0.4394	0.9394	0.0606	0.1200
1.56	0.4406	0.9406	0.0594	0.1182
1.57	0.4418	0.9418	0.0582	0.1163
1.58	0.4429	0.9429	0.0571	0.1145
1.59	0.4441	0.9441	0.0559	0.1127
1.60	0.4452	0.9452	0.0548	0.1109
1.61	0.4463	0.9463	0.0537	0.1092
1.62	0.4474	0.9474	0.0526	0.1074
1.63	0.4484	0.9484	0.0516	0.1057
1.64	0.4495	0.9495	0.0505	0.1040
1.65	0.4505	0.9505	0.0495	0.1023
1.66	0.4515	0.9515	0.0485	0.1006
1.67	0.4525	0.9525	0.0475	0.0989
1.68	0.4535	0.9535	0.0465	0.0973
1.69	0.4545	0.9545	0.0455	0.0957
1.70	0.4554	0.9554	0.0446	0.0940
1.71	0.4564	0.9564	0.0436	0.0925
1.72	0.4573	0.9573	0.0427	0.0909
1.73	0.4582	0.9582	0.0418	0.0893
1.74	0.4591	0.9591	0.0409	0.0878

附录 *A*(续)

标准分 (1) $z=x/\sigma$	面积 (2) 自 $\overline{X}$ 至 z	面积 (3) 较大部分	面积 (4) 较小部分	纵线 (5) y
1.75	0.4599	0.9599	0.0401	0.0863
1.76	0.4608	0.9608	0.0392	0.0848
1.77	0.4616	0.9616	0.0384	0.0833
1.78	0.4625	0.9625	0.0375	0.0818
1.79	0.4633	0.9633	0.0367	0.0894
1.80	0.4641	0.9641	0.0359	0.0790
1.81	0.4649	0.9649	0.0351	0.0775
1.82	0.4656	0.9656	0.0344	0.0761
1.83	0.4664	0.9664	0.0336	0.0748
1.84	0.4671	0.9671	0.0829	0.0734
1.85	0.4678	0.9678	0.0322	0.0721
1.86	0.4686	0.9686	0.0314	0.0707
1.87	0.4693	0.9693	0.0307	0.0694
1.88	0.4699	0.9699	0.0301	0.0681
1.89	0.4706	0.9706	0.0294	0.0669
1.90	0.4713	0.9713	0.0287	0.0656
1.91	0.4719	0.9719	0.0281	0.0644
1.92	0.4726	0.9726	0.0274	0.0632
1.93	0.4732	0.9732	0.0268	0.0620
1.94	0.4738	0.9738	0.0262	0.0603
1.95	0.4744	0.9744	0.0256	0.0596
1.96	0.4750	0.9750	0.0250	0.0584
1.97	0.4756	0.9756	0.0244	0.0573
1.98	0.4761	0.9761	0.0239	0.0562
1.99	0.4767	0.9767	0.0233	0.0551
2.00	0.4772	0.9772	0.0228	0.0540
2.01	0.4778	0.9778	0.0222	0.0529
2.02	0.4783	0.9783	0.0217	0.0519
2.03	0.4788	0.9788	0.0212	0.0508
2.04	0.4793	0.9793	0.0207	0.0498
2.05	0.4798	0.9798	0.0202	0.0488
2.06	0.4803	0.9803	0.0197	0.0478
2.07	0.4808	0.9808	0.0192	0.0468
2.08	0.4812	0.9812	0.0188	0.0459
2.09	0.4817	0.9817	0.0183	0.0449

附录 *A*(续)

标准分 (1) $z=x/\sigma$	面积 (2) 自 $\overline{X}$ 至 z	面积 (3) 较大部分	面积 (4) 较小部分	纵线 (5) y
2.10	0.4821	0.9821	0.0179	0.0440
2.11	0.4826	0.9826	0.0174	0.0431
2.12	0.4830	0.9830	0.0170	0.0422
2.13	0.4834	0.9834	0.0166	0.0413
2.14	0.4838	0.9838	0.0162	0.0404
2.15	0.4842	0.9842	0.0158	0.0396
2.16	0.4846	0.9846	0.0154	0.0387
2.17	0.4850	0.9850	0.0150	0.0379
2.18	0.4854	0.9854	0.0146	0.0371
2.19	0.4857	0.9857	0.0143	0.0363
2.20	0.4861	0.9861	0.0139	0.0355
2.21	0.4864	0.9864	0.0136	0.0347
2.22	0.4868	0.9868	0.0132	0.0339
2.23	0.4371	0.9371	0.0129	0.0332
2.24	0.4875	0.9875	0.0125	0.0325
2.25	0.4878	0.9878	0.0122	0.0317
2.26	0.4881	0.9881	0.0119	0.0310
2.27	0.4884	0.9884	0.0116	0.0303
2.28	0.4887	0.9887	0.0113	0.0297
2.29	0.4890	0.9890	0.0110	0.0290
2.30	0.4893	0.9893	0.0107	0.0283
2.31	0.4896	0.9896	0.0104	0.0277
2.32	0.4893	0.9898	0.0102	0.0270
2.33	0.4901	0.9901	0.0099	0.0264
2.34	0.4904	0.9904	0.0096	0.0258
2.35	0.4906	0.9906	0.0094	0.0252
2.36	0.4909	0.9809	0.0091	0.0246
2.37	0.4911	0.9911	0.0089	0.0241
2.38	0.4913	0.9913	0.0087	0.0235
2.39	0.4916	0.9916	0.0084	0.0229
2.40	0.4918	0.9910	0.0082	0.0224
2.41	0.4920	0.9920	0.0080	0.0219
2.42	0.4922	0.9922	0.0078	0.0213
2.43	0.4925	0.9925	0.0075	0.0203
2.44	0.4927	0.9927	0.0073	0.0203

附录 A(续)

标准分 (1) $z=x/\sigma$	面积			纵线 (5) y
	(2) 自 $\bar{X}$ 至 z	(3) 较大部分	(4) 较小部分	
2.45	0.4929	0.9929	0.0071	0.0198
2.46	0.4931	0.9931	0.0069	0.0194
2.47	0.4932	0.9932	0.0063	0.0189
2.48	0.4934	0.9934	0.0066	0.0184
2.49	0.4936	0.9936	0.0064	0.0180
2.50	0.4938	0.9938	0.0062	0.0175
2.51	0.4940	0.9940	0.0060	0.0171
2.52	0.4941	0.9941	0.0059	0.0167
2.53	0.4943	0.9943	0.0057	0.0163
2.54	0.4945	0.9945	0.0055	0.0158
2.55	0.4946	0.9946	0.0054	0.0154
2.56	0.4948	0.9948	0.0052	0.0151
2.57	0.4949	0.9949	0.0051	0.0147
2.58	0.4951	0.9951	0.0049	0.0143
2.59	0.4952	0.9952	0.0048	0.0139
2.60	0.4953	0.9953	0.0047	0.0136
2.61	0.4955	0.9955	0.0045	0.0132
2.62	0.4956	0.9956	0.0044	0.0129
2.63	0.4957	0.9957	0.0043	0.0126
2.64	0.4959	0.9959	0.0041	0.0122
2.65	0.4960	0.9960	0.0040	0.0119
2.66	0.4961	0.9961	0.0039	0.0116
2.67	0.4962	0.9962	0.0038	0.0113
2.68	0.4963	0.9963	0.0037	0.0110
2.69	0.4964	0.9964	0.0036	0.0107
2.70	0.4965	0.9965	0.0035	0.0104
2.71	0.4966	0.9966	0.0034	0.0101
2.72	0.4967	0.9967	0.0033	0.0009
2.73	0.4968	0.9968	0.0032	0.0096
2.74	0.4969	0.9969	0.0031	0.0093
2.75	0.4970	0.9970	0.0030	0.0091
2.76	0.4971	0.9971	0.0029	0.0088
2.77	0.4972	0.9972	0.0028	0.0936
2.78	0.4973	0.9973	0.0027	0.0081
2.79	0.4974	0.9974	0.0026	0.0081

附录 A(续)

标准分 (1) $z=x/\sigma$	面积			纵线 (5) y
	(2) 自 $\bar{X}$ 至 z	(3) 较大部分	(4) 较小部分	
2.80	0.4974	0.9974	0.0026	0.0079
2.81	0.4975	0.9975	0.0025	0.0077
2.82	0.4976	0.9976	0.0024	0.0075
2.83	0.4977	0.9977	0.0023	0.0073
2.84	0.4977	0.9977	0.0023	0.0071
2.85	0.4978	0.9978	0.0022	0.0069
2.86	0.4979	0.9979	0.0021	0.0067
2.87	0.4979	0.9979	0.0021	0.0065
2.88	0.4980	0.9980	0.0020	0.0063
2.89	0.4981	0.9981	0.0019	0.0061
2.90	0.4981	0.9981	0.0019	0.0060
2.91	0.4932	0.9982	0.0018	0.0058
2.92	0.4982	0.9982	0.0018	0.0056
2.93	0.4983	0.9983	0.0017	0.0055
2.94	0.4984	0.9984	0.0016	0.0053
2.95	0.4984	0.9984	0.0016	0.0051
2.96	0.4985	0.9935	0.0015	0.0050
2.97	0.4985	0.9935	0.0015	0.0048
2.98	0.4986	0.9966	0.0014	0.0047
2.99	0.4986	0.9986	0.0014	0.0046
3.00	0.4987	0.9987	0.0013	0.0044
3.01	0.4987	0.9987	0.0013	0.0043
3.02	0.4987	0.9987	0.0013	0.0042
3.03	0.4938	0.9938	0.0012	0.0040
3.04	0.4938	0.9988	0.0012	0.0039
3.05	0.4939	0.9989	0.0011	0.0038
3.06	0.4939	0.9989	0.0011	0.0037
3.07	0.4939	0.9989	0.0011	0.0036
3.08	0.4990	0.9990	0.0010	0.0035
3.09	0.4990	0.9990	0.0010	0.0034
3.10	0.4990	0.9990	0.0010	0.0033
3.11	0.4991	0.9991	0.0009	0.0032
3.12	0.4991	0.9991	0.0909	0.0031
3.13	0.4991	0.9991	0.0009	0.0030
3.14	0.4992	0.9992	0.0008	0.0029

附录 A(续)

标准分 (1) $z=x/\sigma$	面积			纵线 (5) y
	(2) 自 $\overline{X}$ 至 z	(3) 较大部分	(4) 较小部分	
3.15	0.4092	0.9992	0.0008	0.0028
3.16	0.4992	0.9992	0.0008	0.0027
3.17	0.4992	0.9992	0.0008	0.0026
3.18	0.4993	0.9993	0.0007	0.0025
3.19	0.4993	0.9993	0.0007	0.0025
3.20	0.4993	0.9903	0.0007	0.0024
3.21	0.4993	0.9993	0.0007	0.0023
3.22	0.4994	0.9994	0.0006	0.0022
3.23	0.4994	0.9994	0.0006	0.0022
3.24	0.4994	0.9994	0.0906	0.0021
3.30	0.4995	0.9995	0.0005	0.0017
3.40	0.4997	0.9997	0.0003	0.0012
3.50	0.4998	0.9998	0.0002	0.0009
3.60	0.4998	0.9998	0.0002	0.0006
3.70	0.4999	0.9999	0.0001	0.0004

附录B t概率分布表

自由度 df	单侧概率					
	0.10	0.05	0.025	0.01	0.005	0.0005
	双侧概率					
	0.20	0.10	0.05	0.02	0.01	0.001
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.975	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.941
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.859
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.405
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	6.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.432	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

附录 C　χ^2 概率分布表

df	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
1	0.000157	0.000982	0.000393	0.0458	2.706	3.841	5.024	6.635
2	0.0201	0.0506	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.832	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.9440	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.336	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
23	10.196	11.688	13.091	14.848	32.007	35.172	38.076	41.638
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
25	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
29	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
31	15.655	17.539	19.281	21.434	41.422	44.985	48.232	52.191
32	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486
33	17.073	19.047	20.867	23.110	43.745	47.400	50.725	54.776
34	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061

附录 C(续)

df	0.010	0.025	0.050	0.10	0.90	0.95	0.975	0.99
35	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342
36	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619
37	19.960	22.106	24.075	26.492	48.363	52.192	55.668	59.892
38	20.691	22.878	24.884	27.343	49.513	53.384	56.895	61.162
39	21.426	23.654	25.695	28.196	50.660	54.572	58.120	62.428
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
41	22.906	25.215	27.356	29.907	52.949	56.942	60.561	64.950
42	23.650	25.999	28.144	30.765	54.090	58.124	61.777	66.206
43	24.398	26.785	28.965	31.625	55.230	59.304	62.990	67.459
44	25.148	27.575	29.787	32.487	56.369	60.481	64.201	68.709
45	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957
46	26.657	29.160	31.439	34.215	58.641	62.830	66.617	71.201
47	27.416	29.956	32.268	35.081	59.774	64.001	67.821	72.443
48	28.177	30.755	33.098	35.949	60.907	65.171	69.023	73.683
49	28.941	31.555	33.930	36.818	62.038	66.339	70.222	74.919
50	29.707	32.357	344.764	37.689	63.167	67.505	71.420	76.154

SOURCE: Scientific Tables, 6th ed. (Basel, Switzerland: J. R. Geigy, 1962), p. 36

附录D F概率分布表

1. 显著性水平 $\alpha=0.05$ 时 F 的临界值

分母自由度 df_2	分子自由度 df_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	245.9	248.0	249.1	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

附录 D(续)

2. 显著性水平 $\alpha=0.01$ 时 F 的临界值

分母自由度 df_2	分子自由度 df_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999.5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	98.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	39.64	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.50	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.17	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.10	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.05	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.80	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.34	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.68	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.54	2.45	2.36	2.27	2.17
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81	2.66	2.58	2.50	2.42	2.33	2.23	2.13
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78	2.63	2.55	2.47	2.38	2.29	2.20	2.10
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75	2.60	2.52	2.44	2.35	2.26	2.17	2.06
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73	2.57	2.40	2.41	2.33	2.23	2.14	2.03
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.61	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00